

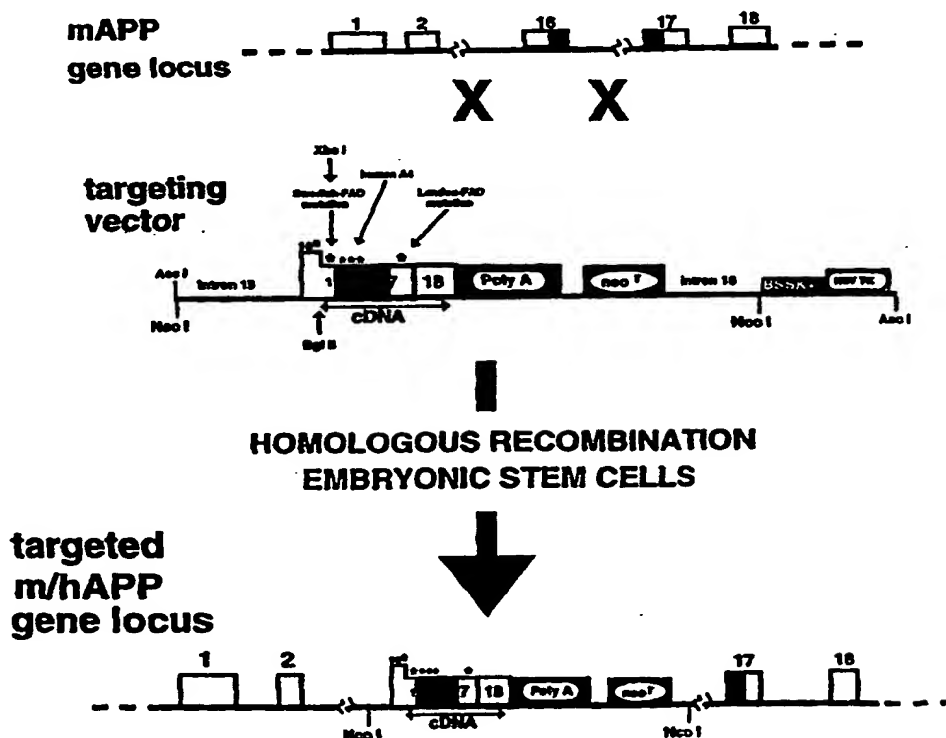
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/00, 15/12, 15/10, 15/90, 5/10, C07K 14/47, A01K 67/027		A1	(11) International Publication Number: WO 99/09150 (43) International Publication Date: 25 February 1999 (25.02.99)
(21) International Application Number: PCT/US97/14507 (22) International Filing Date: 18 August 1997 (18.08.97) (71) Applicant: BAYER CORPORATION [US/US]; 100 Bayer Road, Pittsburgh, PA 15205 (US). (72) Inventor: WIRAK, Dana, Owen; 27 Barberry Lane, Woodbridge, CT 06525 (US). (74) Agents: GREENMAN, Jeffrey, M. et al.; Bayer Corporation and Indiana Corporation, 400 Morgan Lane, West Haven, CT 06516-4175 (US).		(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, RO, RU, SD, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>	

(54) Title: METHOD OF INTRODUCING MODIFICATIONS INTO A GENE

(57) Abstract

The present invention relates to a gene targeting vector and a method of using it to modify nucleic acid sequences. A gene targeting vector in accordance with the invention can comprise: a nucleotide sequence which is effective to achieve homologous recombination at a predefined position of a target gene, operably linked to the 5' terminus of a nucleotide coding sequence which, when inserted into a target gene, codes for at least one amino acid whose identity and/or position is not naturally-occurring in the target gene, and a nucleotide sequence which is effective to achieve homologous recombination at a predefined position of the target gene, operably linked to the 3' terminus of said nucleotide coding sequence. The nucleotide coding sequence can code without interruption for an amino acid sequence, where the amino acid sequence is coded for by two or more exons in a naturally-occurring gene.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

METHOD OF INTRODUCING MODIFICATIONS INTO A GENE

BACKGROUND

The current prior art methods of modifying genes are cumbersome and difficult. For example, mutagenesis of the mouse gene locus via "hit-and-run" and "tag-and-exchange" gene targeting technologies can require the mouse gene locus to be targeted two times in succession using the same ES cell clone. This is a long and laborious process. It is extremely difficult to maintain totipotency of the ES cell through so many manipulations and over such long periods of time in culture. To overcome the difficulties in the prior art, we have developed a novel method of targeting and engineering gene sequences.

The current state of the art provides for three different approaches to the development of transgenic animal models (Lamb, *Nat. Genet.*, 9:4-6, 1995). The first approach utilizes pronuclear injections of recombinant minigenes into the pronuclei of 1-cell embryos. In the second approach, a complete gene residing in yeast artificial chromosomes (YACs), is electroporated into embryonic stem cells (ES cells). The third approach utilizes gene targeting techniques in ES cells to introduce point mutations into a gene present in the ES cell chromosome. The most common approaches to introducing point mutations are "hit-and-run" (Hasty et al., *Nature*, 350:243-246, 1991) or "tag-and-exchange" (Askew et al., *Mol. Cell. Biol.*, 13:4115-4124, 1993), (Stacey et al., *Mol. Cell. Biol.*, 14:1009-1016, 1994) gene targeting procedures.

Recombinant minigenes, when injected into mouse embryos, integrate into the mouse chromosome at random locations. The site of integration can often exert a deleterious influence on the pattern of expression and/or expression level of the recombinant level of the recombinant minigene ("position effect") (Bonnerot et al., *Proc. Natl. Acad. Sci.*, 87:6331-6315,

- 2 -

1990; Brinster et al., *Proc. Natl. Acad. Sci.*, 85:836-840, 1988; Grosveld et al., *Cell*, 51:976-985, 1987).

To illustrate the benefit and ease of the novel compositions, methods, treatments, etc., described herein, we have utilized genes associated with Alzheimer's disease. Therefore, although aspects of this disclosure are written with respect to Alzheimer's diseases, e.g., the APP gene, it is recognized that this invention is in no way limited to such genes and diseases, but may be applied to any nucleic acids, etc., that one desires to target and/or modify.

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by progressive deterioration of memory and cognition. Prominent histopathological features of this disease include the extracellular deposition of amyloid and the accumulation of intracellular neurofibrillary tangles. The principal underlying cellular features of AD are the degeneration affects many types of neurons and may account for the numerous neurological deficits that patients afflicted with the disease encounter. The most notable degeneration occurs in the hippocampus, cerebral cortex, and amygdala, regions of the brain that play a major role in memory, cognition, and behavior.

Although numerous attempts have been made to generate transgenic mouse models for AD via the pronuclear injection approach (Lamb, 1995), only one line of transgenic mice has succeeded in developing extra-cellular plaque-like deposits of beta-amyloid (Games et al., *Nature*, 373:523-527, 1995). This transgenic mouse line utilizes the PDGF promoter to over-express (> 10 fold) human "London"-FAD APP. Because of the artificial nature of the transgene's regulation of gene expression and the aberrant high levels of APP expression, the accumulation of amyloid in this line of transgenic mice may not be fully relevant to the cellular mechanisms involved in Alzheimer's disease.

- 3 -

Two additional papers report only partial success in developing AD-like pathology. In one transgenic model, human APP 751 is over-produced in the brain using the brain-specific enolase promoter (Higgins et al., *Ann. Neurol.*, 35:598-607, 1994). This mouse model exhibits diffuse extra-cellular staining for beta-amyloid, but there was no evidence of accumulations of plaque-like deposits as described by Games et al. (Games et al., 1995). Another transgenic model exhibits intra-cellular deposits of beta-amyloid (La Ferla et al., *Nat. Genet.*, 9:21-30, 1995). This deposition leads to neuropathological processes, including apoptotic neurons and gliosis.

All transgenic mice derived via pronuclear injections retain the ability to express mouse APP. It has been demonstrated that mouse amyloid peptides do not aggregate in solution nearly as well as the human amyloid peptides (Dyrks et al., *FEBS Lett.*, 324:231-36, 1993). It is likely that the mouse amyloid peptide interferes with the process of human amyloid aggregation. This may, in part, explain the necessity in the existing mouse AD model to greatly over-express human amyloid in a mouse brain to develop extra-cellular amyloid deposits.

The human APP gene locus encompasses a very large region (~400 Kb). Transgenic mice have been generated using YACs which appear to contain an intact human APP gene (Lamb et al., 1993; Pearson and Choi, *Proc. Natl. Acad. Sci.*, 30:10578-10582, 1993). But because gene regulatory elements have been identified at considerable distances from the proximal promoter of many genes (e.g., (Grosveld et al., 1987) and (Simonet et al., *J. Biol. Chem.*, 268:8221-8229, 1993)) there is no assurance that a given APP YAC clone contains all necessary APP gene regulatory elements. AD is a complex disease of aging, and the regulation of APP gene expression may play a critical role in the onset and progression of the disease. An accurate mouse model for AD may very well require the presence of critical APP gene regulatory elements which may be missing or altered in the YAC clones. In

- 4 -

addition, the YACs will integrate at random sites in the mouse chromosome after electroporation and expression of the human APP gene may be altered in a detrimental fashion due to "position" effects (see above).

YAC clones are inherently unstable and it can be very difficult to generate transgenic mouse lines where the gene locus resident on the YAC has remained intact. Furthermore, FAD mutations need to be introduced into the very large YACs via homologous recombination in yeast. Determining the stability and integrity of FAD-APP YACs require considerable effort (Lamb et al., 1993, Pearson and Choi, 1993).

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a method of modifying a target nucleic acid. The target nucleic acid preferably comprises a genomic DNA sequence. The invention also relates to recombinant nucleic acid molecules which comprise a nucleotide sequence effective for homologous recombination at a predefined position of a gene and which is operably linked to a nucleotide coding sequence. Such recombinant nucleic acid molecules can be further combined with a vector sequence, a selectable marker, etc., to form a targeting vector useful for modifying a target nucleic acid, e.g., a genomic DNA sequence. The invention also relates to transgenic animals which comprise cells containing a recombinant gene, e.g., an APP gene or a presenilin gene, where the gene has been modified or engineered using the mentioned gene targeting vector. The transgenic animals are useful as animal models for diseases associated with the modified gene locus, e.g., Alzheimer's disease for the APP or presenilin genes.

An object of the invention is a novel gene targeting strategy that facilitates the introduction of one or more specific mutations into any gene in a single double reciprocal homologous recombination step, providing a clear advantage over other gene targeting approaches which use at least two

- 5 -

transfection and screening/selection steps. The gene targeting strategy preferably utilizes double reciprocal homologous recombination and a positive selectable marker gene to facilitate the insertion of gene segments or cDNA's (from the same or a heterologous host) into specific sites within the chromosome of a desired host cell, e.g., an embryonic stem (ES) cell derived from a rodent such as mouse. By the term "cDNA", it is meant a DNA which has been obtained by copying mRNA. The gene segments or cDNA's can be modified to encode one or more mutations. These gene-to-gene segments or gene-to-cDNA fusions, therefore, allow the introduction of one or more specific mutations into the coding sequence of the targeted gene. For some purposes, it may be preferable to employ a cDNA which is modified by the addition of other desired sequences, either coding or non-coding.

An aspect of the invention is a recombinant nucleic acid molecule comprising a nucleotide coding sequence, e.g., a cDNA, which is operably linked at its 5' or 3' terminus, or at both, to a nucleotide sequence which is effective to achieve homologous recombination. The invention also relates to a nucleotide sequence of a rodent APP gene such as a murine APP gene, or other mammal, which is effective to achieve homologous recombination at a predefined position in a target gene, operably linked to the 5' terminus, 3' terminus, or both, of a nucleotide sequence coding for at least one amino acid which is not naturally occurring at a specific amino acid position of the target gene. When the molecule comprises sequences at its 5' and 3' terminus which are homologous to the target gene, the molecule is effective to achieve homologous recombination with the target gene located, e.g., on a chromosome.

The term recombinant means a nucleic acid molecule which has been modified by the hand-of-man, e.g., comprising fragments of nucleic acid from different sources or a nucleic acid molecule from one source which has been engineered. Thus, the nucleic acid molecule is recombinant, e.g.,

- 6 -

because it comprises nucleotide sequences from a rodent (e.g., mouse) gene and a human gene or a synthetic (i.e., engineered) nucleotide sequence.

Homologous recombination is a process in which nucleic acid molecules with similar genetic information line up side by side and exchange nucleotide

5 strands. A nucleotide sequence of the recombinant nucleic acid which is effective to achieve homologous recombination at a predefined position of a target gene therefore indicates a nucleotide sequence which facilitates the exchange of nucleotide strands between the recombinant nucleic acid

10 molecule at a defined position of a target gene, e.g., a mouse APP gene. The effective nucleotide sequence generally comprises a nucleotide sequence which is complementary to a desired target nucleic acid molecule (e.g., the gene locus to be modified), promoting nucleotide base pairing. Any

15 nucleotide sequence can be employed as long as it facilitates homologous recombination at a specific and selected position along the target nucleic acid molecule. Generally, there is an exponential dependence of targeting efficiency on the extent or length of homology between the targeting vector and the target locus. Selection and use of sequences effective for homologous recombination is described, e.g., in Deng and Capecchi, *Mol. Cell. Bio.*, 12:3365-3371, 1992; Bollag et al., *Annu. Rev. Genet.*, 23:199-225, 1989; 20 Waldman and Liskay, *Mol. Cell. Bio.*, 8:5350-5357, 1988.

The nucleotide sequence effective for homologous recombination can be operably linked to a nucleotide sequence, preferably comprising a nucleotide coding sequence, which is to be recombined into the desired target nucleic acid. For example, an aspect of the present invention is to replace all

25 or part of the amino acids comprising exons 16, 17, and 18 of the APP gene with a cDNA coding for all or part of the corresponding amino acids of a human APP gene. This is achieved by attaching a part of the APP gene comprising a part of intron 15 and exon 16 to the 5' terminus of a human cDNA and a part of the APP gene comprising a part of intron 16 to the 3'

- 7 -

terminus of the cDNA to form a targeting vector. The APP gene segments are positioned with respect to the human cDNA in a way such that homologous recombination between them and the mouse gene will result in replacement of exons 16 through 18 with the cDNA. Such positioning, i.e., operable linkage, means that the mouse gene segment is joined to the cDNA whereby the homologous recombination function can be accomplished.

A nucleic acid comprising a nucleotide sequence coding without interruption means that the nucleotide sequence contains an amino acid coding sequence for a polypeptide, with no non-coding nucleotides interrupting or intervening in the coding sequence, e.g., absent intron(s) or the noncoding sequence, as in a cDNA.

An object of the present invention is to introduce modifications into genomic sequences, e.g., by introducing into or replacing a genomic sequence with a cDNA. Such cDNA can comprise one or more mutations, thereby facilitating the introduction of any desired nucleotide sequence into a target nucleic acid. The introduced nucleic acid, e.g., a DNA can particularly encode modifications in, or which span, two or more exons of a desired gene using only a single, double reciprocal homologous recombination event. In one embodiment, two independent point mutations can be introduced into a genomic sequence, where each point mutation is located in a different exon of the same gene. Thus, the coding sequence can be a nucleotide sequence which codes without interruption for an amino acid sequence, where the amino acid sequence is coded for by two or more exons in a naturally-occurring genomic (i.e., gene) sequence. This includes, e.g., a coding sequence for an amino acid sequence which is a cDNA, where the cDNA comprises amino acids coded for by separate exons of a naturally-occurring genomic sequence comprising exons and introns. By the phrase naturally-occurring genomic sequence, it is meant the gene structure as it occurs in nature. For example, a human APP gene contains 18 exons in a

- 8 -

naturally-occurring form which has been described. See, e.g., Yoshikai et al., *Gene*, 87:291-292, 1990. Other gene structures are also possible.

The introduction of point mutations via a replacement type vector has been described (Rubinstein et al., *Nucl. Acid Res.*, 21:2613-2617, 1993).

5 Rubinstein et al. did not consider fusing genomic sequences with cDNA sequences to encode the gene product. Therefore, the gene targeting technology described by Rubinstein et al. would not succeed in introducing the Swedish-London and Swedish-714 stop double mutations into the mouse APP gene locus. The beta-amyloid domain resides on two separate exons
10 (Lemaire et al., *Nucleic Acid Res.*, 17:517-522, 1989; Kang and Muller-Hill, *Biochem. Biophys. Res. Comm.*, 166:1192-2000, 1990). While the Swedish mutation and human amino acid differences reside on exon 16, the London mutation resides on exon 17. Lambda genomic clones are not large enough to encompass both exons (Lamb et al., *Nature Genetics*, 5:22-30, 1993).

15 Therefore, the introduction of the double mutations into a host gene locus (e.g., a mouse APP gene) by the previously described gene-targeting approaches would require multiple gene targeting events utilizing two independent targeting vectors. Thus, it is recognized that in accordance with the present invention mutations which span sequences too large to fit into
20 conventional vectors, targeting strategies, etc. (such as described in Lamb et al., 1993), e.g., two or more exons, can be introduced into genomic DNA by preparing targeting vectors comprising an intron effective for homologous recombination and a contiguous coding sequence, e.g., from the two or more exons.

25 The nucleotide coding sequence can code for at least one amino acid whose identity and/or position is not naturally-occurring in a target gene, e.g., a rodent (e.g., mouse) or non-human mammal gene. This means that the nucleotide coding sequence, when inserted into the target gene such that an open reading frame is formed with the target gene coding sequences,

- 9 -

contains at least one non-identical amino acid from the coding sequence of the unmodified target gene. This can mean amino acid substitution, deletion, or addition. In the examples below which illustrate, but in no way limit the invention, a nucleic acid coding for amino acids of a mouse APP gene are replaced by nucleic acid coding for amino acids of a human APP gene. At least 5 alternative splice forms of APP have been detected (reviewed in Beyreuther et al., *Ann. NY Acad. Sci.*, 695, 91-102 (1993)). The amino acids of a human APP gene means amino acid(s) identified as non-identical when the two APP gene sequences are compared. The amino acid numbering in the patent application refers to the largest alternative splice form of APP which consists of 770 amino acids. See, e.g., Kitaguchi et al., *Nature* 331, 530-532 (1988); Tanaka et al., *Biochem. Biophys. Res. Commun.*, 157, 472-479 (1988). For example, the human amino acid sequence differs in the beta-amyloid domain are at positions 676, 681, and 684. The mouse APP gene contains a glycine at amino acid position 676, and a phenylalanine at amino acid position 681, and an arginine at amino acid position 684. A nucleotide coding sequence, which when inserted into an open reading frame of the mouse APP gene, comprising an arginine at amino acid position 676, a threonine at amino acid position 681, and/or a histidine at amino acid 684 is considered to contain three amino acid(s) whose identify is not naturally-occurring at an amino acid position (i.e., 676, 681, and/or 684) in the target mouse APP gene. See Figure 17 for other differences between the mouse and human APP polypeptide sequence.

A nucleic acid coding for at least one amino acid not naturally occurring in the targeted gene can also comprise, e.g., nucleotides which occur in a naturally-occurring human gene, such as naturally-occurring polymorphisms, alleles, or mutations which are discovered or identified in a natural population. By the term naturally-occurring, it is meant that the nucleic acid is obtained from a natural source, e.g., animal tissue and cells,

- 10 -

body fluids, tissue culture cells, forensic samples. Any other amino acid(s) can be incorporated, as well as, e.g., conservative and non-conservative amino acid substitutions, amino acid(s) obtained from other genes, non-naturally-occurring or engineered sequences, functional and/or selectable coding sequence domains.

In the examples, a mouse APP gene is targeted by the substitution of an amino acid found in a human APP gene. Numerous naturally-occurring mutations have been identified in non-murine APP genes. A nucleic acid according to the present invention can contain such mutations. Other modifications to the sequence can comprise mutations found in familial or genetic cases of disease, preferably Alzheimer's disease, Down's syndrome, or heredity cerebral hemorrhage with amyloidosis Dutch type (HCHWA-D). A nucleotide sequence coding for all or part of an amino acid sequence of a human APP gene can contain codons found in a naturally-occurring gene or transcript, or it can contain degenerate codons coding for the same amino acid sequences.

Preferred human APP amino acid sequences include: Swedish-FAD, KM(670,671)NL; London-FAD, V(717)I; Swedish/London-FAD, KM(670,671)NL, V(717)I; stop codon at position 714; Swedish-FAD, KM(670,671)NL, stop codon at position 714, etc. See Table 1.

An amino acid sequence of a human APP gene comprising a nucleotide sequence to be inserted into a targeted mouse APP gene preferably codes without interruption and comprises arginine at 676, threonine at position 681, histidine at position 684, or combinations thereof, in addition to other mutations and engineered codons.

The present invention also relates to nucleic acids which hybridize to a nucleic acid coding for an amino acid sequence of a human APP gene, preferably under stringent conditions. Such hybridizable sequences are

- 11 -

preferably not a naturally-occurring mouse APP nucleotide sequence;
however, mutant mouse APP sequences can be included.

Hybridization conditions can be chosen to select nucleic acids which have a desired amount of nucleotide complementarity with the nucleotide
5 sequence coding for all or part of an amino acid sequence of a human APP gene. A nucleic acid capable of hybridizing to such sequence, preferably, possesses 50%, more preferably, 70% complementarity, between the sequences. The present invention particularly relates to nucleotide sequences which hybridize to the nucleotide sequence coding for human APP amino
10 acids under stringent conditions. As used here, "stringent conditions" means any conditions in which hybridization will occur where there is at least about 95%, preferably 97%, nucleotide complementarity between the nucleic acids. A nucleotide sequence hybridizing to the coding sequence will have a complementary nucleic acid strand, or act as a template for one in the
15 presence of a polymerase (i.e., an appropriate nucleic acid synthesizing enzyme), which has a corresponding amount of nucleotide identity or similarity. The present invention includes both strands of nucleic acid, e.g., a sense strand and an anti-sense strand. Thus, it is understood that a nucleic acid comprising a nucleotide sequence hybridizing to the coding nucleotide
20 sequence of amino acids of a human APP gene also represents a nucleic acid which possesses at least about 95%, preferably 97% nucleotide sequence identity.

According to the present invention, at least one amino acid not naturally-occurring in the targeted gene also includes amino acids selected
25 from engineered or non-naturally-occurring sequences. In the examples, a mouse APP gene is modified by replacing mouse amino acids with amino acids which naturally occur in a human APP gene. However, the mouse APP gene can also be modified or engineered by the introduction of amino acids which are not based on a human APP gene, e.g., conservative or non-

- 12 -

conservative amino acids, cysteines, prolines, functional and/or selectable domains, etc.

Changes or modifications to the nucleotide coding sequence can be accomplished by any method available, including directed or random mutagenesis to a nucleic acid. These sequence modifications include, e.g., nucleotide substitution which does not affect the amino acid sequence (e.g., different codons for the same amino acid), replacing naturally-occurring amino acids with homologous or conservative amino acids, e.g. (based on the size of the side chain and degree of polarization), small nonpolar: cysteine, proline, alanine, threonine; small polar: serine, glycine, aspartate, asparagine; large polar: glutamate, glutamine, lysine, arginine; intermediate polarity: tyrosine, histidine, tryptophan; large nonpolar: phenylalanine, methionine, leucine, isoleucine, valine. In addition, it may be desired to change the codons in the sequence to optimize the sequence for expression in a desired host.

In addition to a gene segment effective for homologous recombination and coding sequence to be recombined, e.g., a recombinant nucleic acid molecule according to the present invention also can include selection markers, 3' regulatory sequences, regulatory sequences, restriction sites, vector sequences, and sequences and/or modification which enhance homologous recombination.

In order to identify cells which have integrated the nucleic acid molecule, it is desirable to include a selectable marker gene, e.g., neomycin resistance, gene HPRT gene, etc. A selectable marker gene codes for a product which can be directly or indirectly detected in a host in which it is expressed. Selectable marker genes and their use are widely used in molecular biology. When a neomycin resistance gene is utilized, cells having incorporated it can be selected by resistance to G418. A second selectable marker gene can also be incorporated into the vector, e.g., a herpes simplex

- 13 -

virus thymidine kinase gene. Any selectable genes routinely used in host cells can be used in the gene targeting vectors, including HSV TK, neo^r, hygromycin, histidinol, Zeocin (Invitrogen), HPRT, etc. Selectable genes can also be included to select against random integration events. Thus, selection for the first marker (e.g., by positive selection), and absence of the second marker (e.g., by negative selection), permits enrichment for transformed cells containing a modified target nucleic acid sequence, e.g., at the APP gene locus. The choice and arrangement of the selectable marker gene(s) in the recombinant nucleic acid molecule are as the skilled worker would know, e.g., described in U.S. Pat. No. 5,464,764 and Rubinstein et al., *Nucl. Acid Res.*, 21:2613-2617, 1993. A preferred recombinant nucleic acid comprises a selectable marker gene, e.g., a gene for neomycin resistance, in the mouse APP gene segment 3' to the cDNA. The selectable marker genes can be operably linked to regulatory sequences which control their expression, e.g., in a cell or tissue specific manner. Examples of such sequences are described, e.g., in U.S. Pat. No. 5,464,764.

In accordance with the present invention, 3' regulatory nucleotide sequences can be operably linked to a recombinant nucleic acid molecule. For example, it may be desirable to include a transcription termination signal and/or polyadenylation signal (e.g., AATAAA tandem repeat) at the 3' end of the nucleotide sequence to be inserted into the foreign gene. Generally, a selectable marker gene directly follows the transcription, termination and polyadenylations signals. Other sequences can also be included, e.g., nucleotide sequences which regulate the stability of a mRNA.

A recombinant nucleic acid can also comprise nucleotide sequences which affect expression of the gene into which it is combined, e.g., enhancers.

A recombinant nucleic acid molecule according to the present invention can also comprise all or part of a vector. A vector is a nucleic acid

- 14 -

molecule which can replicate autonomously in a host cell, e.g., containing an origin of replication. Vectors can be useful to perform manipulations, to propagate, and/or obtain large quantities of the recombinant molecule in a desired host. A skilled worker can select a vector depending on the purpose desired, e.g., to propagate the recombinant molecule in bacteria, yeast, insect, or mammalian cells. Examples of useful vectors include Bluescript KS+II (Stratagene). The following vectors are provided by way of example, Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pbs., pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16a, pNH18Z, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene), pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other vector, e.g., plasmids, viruses, or parts thereof, may be used as long as they are replicable and viable in the desired host. The vector can also comprise sequences which enable it to replicate in the host whose genome is to be modified. The use of such vector can expand the interaction period during which recombination can occur, increasing the targeting efficiency.

Recombinant nucleic acid molecules according to the present invention can also include sequences and modifications which decrease nonhomologous recombination events and/or enhance homologous recombination. For example, it has been found by Chang & Wislon, *Proc. Natl. Acad. Sci. USA*, 84:4959-63, 1987, that the addition of dideoxy nucleotides to the recessed termini of DNA molecules could enhance homologous recombination 6- to 7-fold relative to nonhomologous events.

Recombinant nucleic acid molecules according to the present invention can be prepared according to the various methods known to the skilled worker in the art, e.g., as mentioned in *Current Protocols in Molecular Biology*, Edited by F.M. Ausubel et al., John Wiley & Sons, Inc; and

Current Protocols in Human Genetics, Edited by Nicholas C. Dracopoli et al., John Wiley & Sons, Inc.

In accordance with the present invention, the novel gene-targeting can be used to modify any desired gene. Figure 15 illustrates several general strategies. Figure 15A shows a "typical" host gene with a DNA sequence consisting of a gene promoter, a series of exons (5 in this example). The exons are depicted as boxes. The gene can contain one or more exons. The line between the boxes (exons) represent the introns. The 5'-end of each intron contains a splice donor site which lies directly juxtaposed to the 3'-nucleotide to the preceding exon. The 3'-end of each intron contains a splice acceptor sequence which lies directly juxtaposed to the 5'-end of the neighboring exon. The 3'-end of the last exon contains a nonsense codon (designated as a stop) to terminate translation. This is followed by 3'-untranslated sequences which are present in the gene transcript and then a transcription termination and polyadenylation signal (designated polyA).

Figure 15B illustrates a targeted gene where a cDNA is inserted directly into an exon (exon 4 in this example) of the gene. Using an appropriately designed gene-targeting construct, any exon of a mouse gene can be targeted in this fashion. This is the approach used in the examples to generate the Swedish and/or London FAD-m/hAPP transgenic mice. The sequence of the cDNA is arranged so that the fusion between the gene and the cDNA creates an "in-frame" sequence that properly encodes the desired protein. The cDNA can be modified to encode one or more mutations (designated at *). The cDNA can be derived from transcripts from other genes of the same species or from genes from other species.

The cDNA is inserted into the mouse by homologous recombination. The recombination occurs between the targeted gene and an exogenously added gene-targeting construct or vector. The vector is preferably linearized. For homologous recombination to insert the cDNA into the proper location

- 16 -

and orientation within the targeted gene, the DNA components of the vector can be arranged in a specific manner. The cDNA is preferably positioned between nucleotide sequences which are homologous to specific locations of the targeted gene.

5 In the gene-targeting vector, there is preferably a gene segment comprising a nucleotide sequence corresponds substantially to an upstream (5'-flanking) region of the targeted gene. This segment comprises contiguous and sufficient upstream (5'-flanking) sequences of the targeted gene to allow efficient recombination to take place, i.e., a nucleotide sequence which is
10 effective for homologous recombination. The segment can be followed by a portion of the targeted gene exon (exon 4 in this example).

 In the gene-targeting vector, the sequence spanning the junction between the 3'-end of the targeted gene exon (exon 4 in this example) and the 5-end of the cDNA are arranged precisely in-frame to conserve the open
15 reading frame to properly encode the desired gene product. In effect, the cDNA and the exon into which it is inserted become the terminal exon of the targeted gene. For proper termination and maturation of the transcript encoded by the targeted gene, transcription termination and polyadenylation signals (designated polyA) are positioned directly after the cDNA (and after
20 the translation of stop codon). Directly following the transcription termination and polyadenylation signals, the gene targeting vector further comprises a selectable marker gene such as the neomycin resistance gene (designated neo^r) or the HPRT gene.

 The gene targeting vector can further comprise a downstream (3'-
25 flanking) region of homology to the targeted gene which is placed directly after the selectable marker gene, e.g., neo^r. The downstream region of homology can comprise contiguous gene sequences but can be any length of sequence providing it is sufficiently long to facilitate homologous recombination. The 5'-end of the downstream region of gene homology can

- 17 -

be located at any position proximal to the targeted gene as long as it lies downstream (3') of the mouse gene sequence which forms the junction between the targeted gene exon and the cDNA. After homologous recombination has taken place, the DNA sequence of the targeted gene positioned between the 3'-end of the upstream region of gene homology and the 5'-end of the downstream region of gene homology will have been deleted.

After homologous recombination takes place, exon sequences lying 5' of the exon/cDNA junction will encode the N-terminal portion of the gene product while the cDNA sequences lying 3' of the exon/cDNA junction will encode the C-terminal portion of the gene product.

Figure 15C illustrates a targeted gene where a cDNA is inserted directly into an intron (intron 3 in this example) of the targeted gene. Using an appropriately gene-targeting construct, any intron of a gene could be targeted in this fashion.

The sequence of the cDNA is arranged so that it functions as the terminal exon of the targeted gene. To form an open reading frame between the targeted and human coding sequence, the codon reading-frame of the cDNA sequence is positioned in-frame with the codon reading-frame of the nearest upstream (5') exon (exon 3 in this example). For proper splicing of messenger RNA to occur, a functional splice acceptor site immediately preceding (5') the cDNA can be included. Thus, after splicing of the exon with the cDNA, a resultant transcript from the targeted gene will encode the desired gene product. As mentioned above, a cDNA from various sources can be utilized and it can be modified to encode mutations. The arrangement of the gene targeting vector is as described above.

Figure 15D illustrates a targeted gene where a gene segment from another the same or different species (designated as foreign gene segment) is inserted directly into an intron (intron 3 in this example) of the targeted gene.

- 18 -

Using an appropriately gene-targeting construct, any intron of a gene can be targeted in this fashion. In this example, the sequence of the foreign gene segment contains normal exons and introns from another gene. The sequences of the gene-targeting construct are arranged such that the foreign gene segment functions as the terminal set of exons for the targeted gene. The codon reading-frame of the exons of the foreign gene segment can be arranged in-frame with the codon reading-frame of the nearest upstream (5') exon (exon 3 in this example) to form a complete open-reading frame. For proper splicing of messenger RNA to occur, a functional splice acceptor site immediately preceding the 5' exon of the foreign gene segment can be included. Thus, after splicing of the exon with the exons of the foreign gene segment, the transcript from the targeted gene will encode the desired gene product. The foreign gene segment can be obtained from various sources, as desired, and can be engineered to encode one or more mutations. The arrangement of a gene targeting vector is described above.

Another aspect of the present invention relates to host cells comprising a recombinant nucleic acid of the invention. A cell into which a nucleic acid is introduced is a transformed cell. Host cells include, mammalian cells, e.g., rodent, murine Ltk-, murine embryonic stem cells, COS-7, CHO, HeLa, insect cells, such as Sf9 and Drosophila, bacteria, such as E. coli, Streptococcus, bacillus, yeast, fungal cells, plants, embryonic stem cells (e.g., mammalian, such as mouse or human), neuronal cells (primary or immortalized), e.g., NT-2, NT-2N, PC-12, SY-5Y, neuroblastoma. See, also *Methods in Enzymology*, Volume 185, ed., D.V. Goeddel. A nucleic acid can be introduced into the cell by any effective method including, e.g., calcium phosphate precipitation, electroporation, injection, DEAE-Dextran mediated transfection, fusion with liposomes, and viral transfection. When the recombinant nucleic acid is present in a host cell, it is preferably

- 19 -

integrated by homologous recombination into a chromosome residing in the host cell.

The present invention also relates to a recombinant nucleic acid coding for a recombinant polypeptide, which nucleic acid is a product of the gene which has been modified by the gene targeting vector. A gene can code for different nucleic acid transcripts, depending on splicing, where it is expressed, etc. All such nucleic acids are a product of the recombinant gene and thus relate to the present invention. Such nucleic acids can code for recombinant polypeptides which are also an object of the present invention.

The recombinant polypeptides can be used, e.g., as antigens to generate specific antibodies as diagnostic, research, and therapeutic tools.

A recombinant nucleic acid and a recombinant polypeptide can incorporate at least one amino acid or coding sequence thereof from a heterologous species. If, e.g., a non-human mammal sequence contains at least one amino acid of a human sequence, the modified sequence is described as "humanized." By "humanized" it is meant, e.g., a mouse polypeptide containing one or more amino acids which are present in the human polypeptide (and which differ from the amino acids present in the mouse gene).

Thus, in the examples, humanized mouse APP nucleic acids and polypeptides were created by substituting a human amino acid for a mouse amino acid at corresponding locations. A recombinant nucleic acid can be an unprocessed RNA transcript comprising introns or it can comprise a nucleotide sequence coding without interruption for amino acids, e.g., where the nucleic acid is a modified APP gene, it can code for amino acids 1-770, 1-713, 1-751, and 1-695. For example, a nucleic acid coding for a recombinant APP polypeptide can be a transcript from an APP gene modified in accordance with the present invention, e.g., by homologous recombination with a human cDNA and a mouse gene. The recombinant nucleic acid can

- 20 -

comprise mutations in the APP gene, e.g., Swedish-FAD, London-FAD, etc., as described above.

The present invention also relates to a non-human transgenic animal, preferably a mammal, more preferably a rodent such as a mouse, which
5 comprises a gene, which has been engineered employing a recombinant nucleic acid according to the present invention. Generally, a transformed host cell, preferably a totipotent cell, whose endogenous gene has been modified using a recombinant nucleic acid as described above is employed as a starting material for a transgenic embryo. The preferred methodology for
10 constructing such a transgenic embryo involves transformed embryonic stem (ES) cells employing a targeting vector comprising a recombinant nucleic acid according to the invention. A particular gene locus, e.g., APP, is modified by targeted homologous recombination in cultured ES cells employing a targeting vector comprising a recombinant nucleic acid according to the
15 invention. The ES cells are cultured under conditions effective for homologous recombination. Effective conditions include any culture conditions which are suitable for achieving homologous recombination with the host cell chromosome, including effective temperatures, pH, medias, additives to the media in which the host cell is cultured (e.g., for selection,
20 such as G418 and/or FIAU), cell densities, amounts of DNA, culture dishes, etc. Cells having integrated the targeting vector are selected by the appropriate marker gene present in the vector. After homologous recombination has been accomplished, the cells contain a chromosome having a recombinant gene. In a preferred embodiment, this recombinant gene
25 contains host genomic sequences (e.g., mouse) fused to a donor cDNA (e.g., human). The cDNA can contain multiple mutations, etc., which are not naturally-occurring in the target gene. No further gene engineering steps are necessary. Thus, in accordance with the present invention, a single step has resulted in a modified gene containing as many modified sequences as

- 21 -

desired. Another aspect of the present invention involves employing a cDNA with sufficient nucleotide sequence dissimilarity between it and the native target gene sequence to avoid inappropriate intra-recombination and inter-recombination events, subsequent to the first gene targeting step.

5 The transformed or genetically modified cells can be used to generate transgenic non-human mammals, e.g., rodents (such as mice or rats), by injection into blastocysts and allowing the chimeric blastocysts to mature, following transfer into a pseudopregnant mother. See, e.g.,
10 *Teratomacarcinoma and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed., IRL Press. Various stem cells can be used, as known in the art, e.g., AB-1, HM-1 D3, CC1.2, E-14T62a, preferably ES cell line G1 derived from inbred mouse strain 129/SvEvT.

 In accordance with the present invention, a transformed cell contains a recombinant gene integrated into its chromosome at the targeted gene locus.
15 A targeting vector which comprises sequences effective for homologous recombination at a particular gene locus, when introduced into a cell under appropriate conditions, will recombine with the homologous sequences at the gene locus, introducing a desired gene segment (e.g., a cDNA) into it. When recombination occurs such that insertion results, the nucleic acid is integrated
20 into the gene locus. The gene locus can be the chromosomal locus which is characteristic of the species, or it can be a different locus, e.g., translocated to a different chromosomal position, on a supernumerary chromosome, on an engineered "chromosome," etc. In the examples described below, the sequences of the human APP gene are integrated by homologous
25 recombination into the normal APP gene loci on murine chromosome 16. By recombinant, it is meant that the nucleotide sequences come from different sources, e.g., mouse and human.

 A transgenic non-human mammal comprising a recombinant gene, which when mutant results in Alzheimer's disease, can express the gene in an

- 22 -

amount effective to produce neuronal cell degeneration and/or apoptosis. The gene can also be expressed in an amount effective to cause a behavioral or cognitive dysfunction, wherein the dysfunction is conferred by the recombinant gene. Such gene can be, e.g., PS1, PS2, S182 (e.g.,
5 Sherrington et al., *Nature*, 375:754-760, 1995), STM2, E5-1, apolipoprotein E, apoptosis genes such as ALG-1 to -6 (Vito et al., *Science*, 271:521, 1995), Bcl-2/Bax gene family, etc.

A transgenic non-human animal according to the present invention can comprise one or more genes which have been modified by genetic
10 engineering. For example, a transgenic animal comprising an APP gene which has been modified by targeted homologous recombination in accordance with the present invention can comprise other mutations, including modifications at other gene loci and/or transgenes, including PS1, PS2, S182 (e.g., Sherrington et al., *Nature*, 375:754-760, 1995), STM2, E5-
15 1, apolipoprotein E, apoptosis genes such as ALG-1 to -6 (Vito et al., *Science*, 271: 521, 1995), Bcl-2/Bax gene family, etc. Modifications to these gene loci and/or introduction of transgenes can be accomplished in accordance with the methods of the present invention, or other methods as the skilled worker would know, e.g., by pronuclear injection of recombinant genes into
20 pronuclei of one-cell embryos, incorporating an artificial yeast chromosome into embryonic stem cells, gene targeting methods, embryonic stem cell methodology. See, e.g., U.S. Patent Nos. 4,736,866; 4,873,191; 4,873,316; 5,082,779; 5,304,489; 5,174,986; 5,175,384; 5,175,385; 5,221,778; Gordon et al., *Proc. Natl. Acad. Sci.*, 77:7380-7384 (1980); Palmiter et al., *Cell*,
25 41:343-345 (1985); Palmiter et al., *Ann. Rev. Genet.*, 20:465-499 (1986); Askew et al., *Mol. Cell. Bio.*, 13:4115-4124 (1993); Games et al. *Nature*, 373:523-527 (1995); Valancius and Smithies, *Mol. Cell. Bio.*, 11:1402-1408 (1991); Stacey et al., *Mol. Cell. Bio.*, 14:1009-1016 (1994); Hasty et al.,

Nature, 350:243-246 (1995); Rubinstein et al., *Nucl. Acid Res.*, 21:2613-2617 (1993).

A recombinant nucleic acid molecule according to the present invention can be introduced into any non-human mammal, including a rodent, mouse (Hogan et al., *Manipulating the Mouse Embryo: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1986), rat, pig (Hammer et al., *Nature*, 315:343-345, 1985), sheep (Hammer et al., *Nature*, 315:343-345, 1985), cattle or primate. See also, e.g., Church, *Trends in Biotech.* 5:13-19, 1987; Clark et al., *Trends in Biotech.* 5:20-24, 1987; and DePamphilis et al., *BioTechniques*, 6:662-680, 1988.

A transgenic non-human animal and a recombinant nucleic acid molecule according to the present invention is useful as described in U.S. Pat. Nos. 5,304,489, 5,221,778, 5,175,385, 5,175,384, 5,175,383, 5,087,571, 5,082,779, 4,736,866, 4,873,191, and other transgenic animal patents. For example, a recombinant nucleic acid molecule comprising a coding sequence for at least one amino acid of a human APP gene is useful as a hybridization probe for detecting and diagnosing Alzheimer's disease, e.g., nucleotide variations and genetic polymorphisms present in a nucleic acid can be detected in accordance with various methods, e.g., U.S. Pat. 5,468,613; Conner et al., *Proc. Natl. Acad. Sci.* 80, 78 (1983); Angelini et al., *Proc. Natl. Acad.*, 83, 4489 (1986); Myers et al., *Science*, 230, 1242 (1985). The nucleic acid can also be operably linked to an expression control sequence to produce polypeptide encoded by it. The operable linkage of nucleic acid and expression control sequence can be introduced into a desired host, and cultured under conditions effective to achieve expression of a polypeptide coded for the nucleic acid. An expression control sequence is similarly selected for host compatibility and a desired purpose, e.g., high copy number, high amounts, induction, amplification, controlled expression. Other sequences which can be employed, include, enhancers such as from

- 24 -

SV40, CMV, inducible promoters, neuronal specific elements, or sequences which allow selective or specific cell expression, such as in neuronal cells, glial cells, etc. The expression control sequence includes mRNA-related elements and protein-related elements. Such elements include promoters, enhancers (viral or cellular), ribosome binding sequences, transcriptional terminators, etc. An expression control sequence is operably linked to a nucleotide coding sequence when the expression control sequence is positioned in such a manner to effect or achieve expression of the coding sequence. For example, when a promoter is operably linked 5' to a coding sequence, expression of the coding sequence is driven by the promoter. The resulting polypeptides can be used to generate antibodies for diagnostic purposes, etc. The operable linkage with an expression control sequence can also occur *in situ* as a result of homologous recombination at the desired gene locus, e.g., a mouse APP gene.

A further aspect is the expression of a modified mRNA and polypeptides encoded by a recombinant nucleic acid molecule of the present invention in a transgenic animal, preferably a non-human mammal, as a model for diseases associated with the gene, e.g., the APP, PS1, and PS2 genes with Alzheimer's disease (AD), Down's syndrome, and heredity cerebral hemorrhage with amyloidosis Dutch type (HCHWA-D). Expression of a modified gene product in a transgenic non-human mammal and its consequent phenotype can therefore be used as a model for diseases and pathologies, e.g., as an AD model for genes associated with Alzheimer's disease. As described in the examples below, a mouse APP gene is modified by the introduction of mutations which are associated with an Alzheimer's phenotype in humans. Transgenic mice comprising cells which contain such a modified or recombinant APP gene can be used to design therapies. For example, active agents, e.g., synthetic, organic, inorganic, or nucleic acids based molecules, can be administered to a transgenic non-human mammal

- 25 -

according to the present invention to identify agents which either inhibit, prevent, and/or reduce the appearance of an A β peptide in the brain, the AD pathology, neurodegeneration, apoptosis, cognitive deficits, and/or behavioral symptoms, etc. Thus, another aspect of the invention is to provide a method to assist in the advancement of the treatment and/or prevention of the
5 aforementioned symptoms (e.g., neurodegeneration or apoptosis) caused by the APP gene, or a fragment thereof. Other genes and therapies can be used analogously.

Such a mammal model can also be used to assay for agents, e.g., zinc,
10 and factors, e.g., environmental, which exacerbate and/or accelerate the diseases. See, e.g., Bush et al., *Science*, 265:1464-1467, 1994. A transgenic non-human animal can also be useful as pets, food sources (e.g., mice for snakes), in toxicity studies, etc.

Moreover, a non-human mammal containing a recombinant nucleic
15 acid according to the present invention can be used in a method of screening a compound for its effect on a phenotype of a mammal, preferably a mouse, where the phenotype is conferred by the recombinant nucleic acid. By "phenotype," it is meant, e.g., a collection of morphological, physiological, biochemical, and behavioral traits possessed by a cell or organism that results
20 from the interaction of the genotype and the environment. A phenotype can be behavioral, e.g., occurrence of seizures or cognitive performance, or it can be physiological and/or pathological, e.g., occurrence of neuronal cell degeneration, neuronal cell apoptosis, accumulation of A β peptide in the brain of the mammal, altered carboxy-terminal processing of the APP
25 polypeptide, etc. According to such a method of detection, a compound can be administered to a mammal containing a modified APP gene and then the existence of an effect on the phenotype of the mammal can be determined. Observation can be accomplished by any means, depending on the specific phenotype which is being examined. For example, the ability of a test

- 26 -

compound to suppress a behavioral phenotype can be detected by measuring the latter phenotype before and after administration of the test compound.

The invention also relates to a transgenic non-human mammal comprising cells that contain a recombinant gene modified by a gene targeting vector. For example, such recombinant gene or nucleic acid can code for a humanized mouse polypeptide comprising at least one amino acid coded for by a human gene, e.g., where the gene is the APP, PS1, or PS2 gene. In the case of the APP gene, the gene can code for, e.g., amino acids 1-665 of a mouse APP gene and amino acids 666-770 of a human APP gene, and having a phenotype conferred by the modified gene, e.g., accumulation of A β peptide or other related peptide in the brain, abnormal processing of the APP polypeptide, etc.

The level of expression of the recombinant gene can be any amount which can produce a phenotype in the non-human mammal, which phenotype can be distinguished from mammals which do not possess the modified gene locus, i.e., a control mammal, e.g., an amount effective to produce neuronal cell degeneration and/or apoptosis and/or an amount effective to cause a behavioral and/or cognitive effect or dysfunction where the gene is an alzheimer's disease associate gene.

A non-human mammal containing a modified APP gene can also be characterized by accumulation of the A β peptide in its brain. The accumulation can be in any quantity which is greater than that observed in mammals not containing the modified gene locus. The phenotype conferred by the modified APP gene can occur before or after accumulation can be detected. The expression and/or accumulation of the APP polypeptide, and its processed derivatives, and the nucleic acids which encode it, can be measured conventionally, e.g., by immunoassay or nucleic acid hybridization, either *in situ* or from nucleic acid isolated from host tissues.

- 27 -

The identification of agents which prevent and/or treat symptoms associated with expression of the modified gene can be determined routinely. For example, an active agent can be administered to a transgenic mammal comprising a modified gene according to the present invention and then its effect on a behavior or pathology, e.g., A β deposition in the brain, apoptosis, and/or neurodegeneration, can be determined. The agent can be administered acutely (e.g., once or twice) or chronically by any desired route, e.g., subcutaneously, intravenously, transdermally, or intracathically. The formulation of the agent is conventional, see, e.g., *Remington's Pharmaceutical Sciences*, Eighteenth Edition, Mack Publishing Company, 1990. In a test, e.g., an agent can be administered in different doses to separate groups of transgenic mammals to establish a dose-response curve to select an effective amount of the active agent. Such effective amount can be extrapolated to other mammals, including humans.

The transgenic mammal, preferably a mouse, according to the present invention therefore permits the testing of a wide variety of agents and therapies. In AD, for example, a number of different agents have been identified which affect the cognitive dysfunction associated with the diseases, e.g., cholinergic agents, such as muscarine agonists, acetylcholinesterase inhibitors, acetylcholine precursors, biogenic amines, nootropics, angiotensin converting enzyme (ACE), and vitamin E. In addition, agents which regulate APP or A β expression, A β deposition, and physiological changes associated with A β expression and deposition can also be identified, e.g., calcium homeostasis, inflammation, neurofibrillary tangles. See, e.g., Pavia et al., *Annual Reports of Medicinal Chemistry*, 25:2129, 1989; John et al., *Annual Reports of Medicinal Chemistry*, 28:197-203, 1993. Additionally, active agents which block apoptosis, e.g., free radical scavengers, such as glutathionines, can be administered. Such effects on AD can be assayed in either behavioral or physiological and/or histological studies.

- 28 -

For example, spatial learning and memory abilities in mice can be tested in a Morris water maze. See, e.g., Yamaguchi et al., *NeuroReport*, Vol. 2, 781-784 (1991). Additionally, other behavioral tests can be used, e.g., Swim Test, Morris et al., *Learning and Motivation*, 12, 239-260, 1981; Open-field test, Knardahl et al., *Behav. Neurol. Biol.* 27, 187-200, 1979; and tests and models used routinely, e.g., in mice, rats, and other rodents.

According to the present invention, differences in, e.g., levels of expression, cellular localization, and/or onset of expression of the recombinant gene can be used to model a disease, e.g., AD and other diseases associated with APP expression and the differing stages and progressions of the disease, e.g., cell degeneration, cell death, astrogliosis, and/or amyloidosis. Having a range of expression phenotypes can be useful to identify different therapies and drug treatments and also diagnostically to identify a disease's progression. For example, the specific treatments can depend on the region of the brain in which an APP peptide is expressed, how much of it is expressed, and its temporal progression of expression. Thus, mammals having different phenotypes can be used as models for determining therapies which are selective for different stages of the disease and for studying disease progression and intervention.

DESCRIPTION OF THE FIGURES

Figure 1. Schematic of p35A; mouse APP exon 16 genomic clone

The ~15 Kb Not I genomic fragment (shown) was isolated from the lambda clone 35A and cloned into the Not I site of Bluescript II SK+. Exon 16 is indicated and begins approximately 9.5 Kb from the 5'-end of the genomic fragment. The indicated restriction enzyme recognition sites were placed for reference.

Figure 2. Restriction Map pMTI-2396

pMTI-2396 contains mouse APP exon 16 and was derived from the ~5.5 Kb NcoI fragment from p35A (NcoI at position 7645 to NcoI at position 13176, Figure 1). The 5.5 Kb NcoI fragment was inserted into NcoI-modified Bluescript II SK+ at the NcoI site. All recognition sites for the indicated restriction enzymes are designated. Sequence from positions 29 to 5560 were derived from the mouse APP gene and the remaining sequences were derived from Bluescript II SK+.

Figure 3. Restriction map of pRA3

pRA3 contains mouse APP intron 15 sequences and was derived from the ~3 Kb NcoI fragment from p35A (NcoI at position 4816 to NcoI at position 7645, Figure 1). The 3 Kb fragment was inserted into NcoI-modified Bluescript II SK+ at the NcoI site. All recognition sites for the indicated restriction enzymes are designated. Sequence from positions 29 to 2858 were derived from the mouse APP gene and the remaining sequences were derived from Bluescript II SK+.

Figure 4. Restriction map of pN2C4

- 30 -

pN2C4 contains mouse APP intron 16 sequences and was derived from the ~1.9 Kb NcoI fragment from p35A (NcoI at position 13176 to NcoI at position 14992, Figure 1). The ~1.9 Kb fragment was inserted into NcoI-modified Bluescript II SK+ at the NcoI site. All recognition sites for the indicated restriction enzymes are designated. Sequence from positions 29 to 1845 were derived from the mouse APP gene and the remaining sequences were derived from Bluescript II SK+.

Figure 5. Restriction map of pMTI-2398; Swedish-FAD targeting vector

The mouse APP intron 15 and exon 16 sequences encompass positions 30 to 1960 (BglII site). The human APP cDNA and genomic polyadenylation sequences are contained in sequences between positions 1960 and ~4556. The neomycin resistance gene lies between positions ~4556 and ~6460. Mouse APP intron 16 sequences are contained between positions ~6460 and 9872. The Bluescript II SK+ sequences are between positions ~9872 and ~30. All recognition sites for the indicated restriction enzymes are designated.

Figure 6. Restriction map of pMTI-2453; London-FAD targeting vector

The HSV TK gene is located between positions ~17 and ~2893. The mouse APP intron 15 and exon 16 sequences encompass positions ~2906 to 4835 (BglII site). The human APP cDNA and genomic polyadenylation sequences are contained in sequences between positions 4835 and ~7452. The neomycin resistance gene lies between positions ~7452 and ~9323. Mouse APP intron 16 sequences are contained between positions ~9323 and 12750. The Bluescript SK+II sequences are between positions ~12750 and ~37. All recognition sites for the indicated restriction enzymes are designated.

- 31 -

Figure 7. Restriction map of pMTI-2454; Swedish/London-FAD targeting vector

The HSV TK gene is located between positions ~17 and ~2893. The mouse APP intron 15 and exon 16 sequences encompass positions ~2906 to 4835 (BglII site). The human APP cDNA and genomic polyadenylation sequences are contained in sequences between positions 4835 and ~7452. The neomycin resistance gene lies between positions ~7452 and ~9323. Mouse APP intron 16 sequences are contained between positions ~9323 and 12750. The Bluescript II SK+ sequences are between positions ~12750 and ~37. All recognition sites for the indicated restriction enzymes are designated.

Figure 8. Restriction map of pMTI-2455 (Swedish-FAD APP713 targeting vector)

The HSV TK gene is located between positions ~17 and ~2893. The mouse APP intron 15 and exon 16 sequences encompass positions ~2906 to 4835 (BglII site). The human APP cDNA and genomic polyadenylation sequences are contained in sequences between positions 4835 and ~7452. The neomycin resistance gene lies between positions ~7452 and ~9323. Mouse APP intron 16 sequences are contained between positions ~9323 and 12750. The Bluescript II SK+ sequences are between positions ~12750 and ~37. All recognition sites for the indicated restriction enzymes are designated.

Figure 9. Oligonucleotides

Oligonucleotides are designated in the 5' to 3' direction.

Figure 10. Schematic outline of m/hAPP gene products produced in transgenic mouse lines.

- 32 -

The protein m/hAPP exhibits amino acid sequence identity with mouse APP with the exception of those residues indicated by (asterisks, see text above). m/hAPP protein spans the membrane once as indicated. The bA4 peptide region (indicated by red) partially resides in the transmembrane and extracellular domains. The APP751 alternative splice form of APP has the 56 amino acid Kunitz protease inhibitor domain while the APP770 splice form of the protein has both the Kunitz and the 19 amino acid OX domains. The APP695 alternative splice form of APP contains neither Kunitz nor OX domains. Other splice forms are not indicated. There are two possible N-linked glycosylation sites (CHO) in the extracellular domain of APP. A highly negatively-charged domain and a cysteine-rich domain are symbolized by a minus sign and S-S bridges respectively. The signal peptide (SP) is located at the N-terminus (see Unterbeck et al.).

Figure 11. Gene-targeting strategy: Construction of targeting vectors.

The schematic of the Nco I APP gene fragment represents the ~5.5 Kb NcoI mouse APP gene fragment in pMTI-2396 (Figure 2). The regions indicated in red represent the coding sequences for mouse b-amyloid domain. The schematic for the targeting vector represents the linearized (using AscI) DNA from clone pMTI-2454 (Figure 7). The targeting vectors for pMTI-2453 (Figure 6) and pMTI-2455 (Figure 8) are identical to pMTI-2454 with the exception of the FAD mutation and the orientation of the HSV TK gene (see text). pMTI-2398 is similar to pMTI-2454 with the exception of FAD mutation and the absence of the HSV TK gene (see text). The FAD mutations are indicated by black asterisks and the mutations to "humanize" the b-amyloid domain are indicated by green asterisks. The neomycin resistance gene is designated by neo^r and Bluescript II SK+ sequences are designated by BSSK+.

- 33 -

Figure 12. Gene-targeting strategy: Homologous recombination.

The linearized targeting vector (Figure 11) was electroporated into ES cells. Homologous recombination occurred between mouse APP sequences contained in the targeting vector and mouse APP genomic sequences on chromosome 16. The resulting targeted m/hAPP gene locus is schematically shown. The FAD mutations are indicated by asterisks and the mutations to "humanize" the b-amyloid domain are indicated by asterisks.

Figure 13. Gene-targeting strategy: Targeted m/hAPP gene locus.

The comparison of the mouse APP and targeted m/hAPP gene loci is shown schematically. In normal mouse, the b-amyloid, transmembrane, and cytoplasmic domains of APP are encoded by mouse APP exons 16, 17, and 18. In the case of the targeted m/hAPP gene locus, the b-amyloid, transmembrane, and cytoplasmic domains of m/hAPP are encoded by human cDNA sequences. The remainder of m/hAPP is encoded by mouse APP exons 1 through 15. The FAD mutations are indicated by asterisks and the mutations to "humanize" the b-amyloid domain are indicated by asterisks.

Figure 14. Strategy for Southern-blot detection of ES cells having a targeted m/hAPP gene locus containing the Swedish-FAD mutation (e.g.; transgenic lines ES5007 and ES5103).

The schematics for the mouse and m/hAPP loci are indicated. The restriction enzymes XbaI and NcoI are designated by X and N respectively. The box represents human APP cDNA and genomic sequences while the box represents the neomycin resistance gene.

Figure 15. Gene Targeting Strategies

A. Normal Gene

- 34 -

B. Targeted gene. Fusion of a gene with cDNA (in-frame fusion of mouse exon sequences with cDNA). * represents one or more mutations.

5 C. Targeted gene. Fusion of a target gene with cDNA (cDNA is inserted into a mouse intron (intron 3 for example). The cDNA is directly preceded by a splice acceptor site. The sequence of the insert is formatted so that splicing of the 3'-sequence of the exon (exon 3 for example) with the 5'-sequence of the cDNA will create a mature transcript encoding the appropriate gene product). * represents one or more mutations.

10 D. Targeted gene. Fusion of a targeted gene with a foreign (same or different species) gene segment including one or more exons inserted into the intron of the targeted gene. The sequence of the insert is formatted so that splicing of the 3'-sequence of the mouse exon (exon 3 for example) with the 5'-sequence of the other mouse gene or species exon (exon 4' for example) will create a mature transcript encoding the appropriate gene product). *
15 represents one or more mutations.

Figure 16. Amino acid sequence of human APP.

Figure 17. Sequence of mouse exon 16 locus

Figure 18. Sequence of pMTI-2398 (Swedish-FAD APP targeting vector)

Figure 19. Sequence of pMTI-2453 (London-FAD APP targeting vector)

20 Figure 20. Sequence of pMTI-2454 (Swedish/London-FAD APP targeting vector)

Figure 21. Sequence of pMTI-2455 (Swedish-FAD APP713 targeting vector)

- 35 -

Figure 22. Sequence of APP genomic clone containing human APP polyadenylation signals.

- 36 -

EXAMPLES

Four independent lines of transgenic mice (lines ES5007, ES5103, ES5401 and ES5403) have been created via a novel gene targeting technique applied to embryonic stem cells. In each line, the mouse APP gene has been modified to encode a mouse/human hybrid APP (m/hAPP) where amino acid residues 666-770 of APP770 are now encoded by human cDNA sequences instead of mouse genomic exons (exons 16, 17, and 18). Within these residues only three amino acid differences exist between the mouse and human proteins (Gly(676) to Arg, Phe(681) to Thr, and Arg(684) to His). This exon-cDNA fusion gene, therefore, encodes an APP containing a "humanized" beta-amyloid domain (aa residues 672 to 712).

In each transgenic mouse line, the human cDNA sequences have been modified to introduce one or more mutations proximal to the "humanized" beta-amyloid domain. In transgenic mouse line ES5007, m/hAPP has been mutated to include the "Swedish"-FAD mutation (KM to NL, positions 670 and 671)(Cai et al., 1993, Citron et al., 1994). Transgenic mouse lines ES5401 and ES5403 encode m/hAPP which have been mutated to include the "London"-FAD mutation (V to I, position 717) (Suzuki et al., 1994, Gravina, 1995). Transgenic mouse line ES5103 encodes m/hAPP which has been mutated to include both "London" and "Swedish" FAD mutations. A fifth transgenic mouse line ES5215 can be produced which encodes m/hAPP that has been mutated to include both the "Swedish" FAD mutation and a premature stop codon (T to stop at position 714). With the exception of the changes mentioned above, the remainder of the m/hAPP sequences are identical to those found in normal mouse APP.

We have shown that the targeted Swedish-FAD m/hAPP and Swedish/London-FAD m/hAPP genes express m/hAPP protein at levels approaching those observed for mouse APP in brain.

- 37 -

Notably, we have observed that the Swedish FAD mutation alters significantly the proteolytic processing of APP resulting in differences in the appearance of C-terminal fragments. The observed changes in processing is consistent with the Swedish-FAD mutation inducing the beta-secretase cleavage site to be utilized predominately over the alpha-secretase cleavage site as previously observed in cell culture experiments (see below).

Messenger RNA from the Swedish-FAD m/hAPP gene was found be abundantly expressed in the brain from homozygous ES5007 mice as well. The amount of Swedish-FAD m/hAPP mRNA in homozygous ES5007 brain was determined to be approximately 55% of the mAPP mRNA levels observed in control mouse brain. In concordance, the APP mRNA levels in heterozygous ES5007 mouse brain were found to be approximately 75% of the level observed in control mouse brain.

The reverse transcriptase-PCR (rtPCR) technique was used to identify mouse APP and Swedish-FAD m/hAPP transcripts in mouse brain. Homozygous ES5007 mice were found to express mRNA exclusively from the targeted Swedish-FAD m/hAPP gene. No mRNA species containing sequences from mouse APP exons 16, 17, or 18 was detected in homozygotes. As would be expected, heterozygous ES5007 mice were found to express mRNA transcripts from both normal mouse and Swedish-FAD APP alleles.

Western-blot analyses have demonstrated that Swedish-FAD m/hAPP and Swedish/London-FAD protein is expressed in the brain of ES5007 and ES5130 mice, respectively. Swedish-FAD m/hAPP protein is expressed in the brain of homozygous ES5007 mice at approximately 87% of the level observed for mouse APP in non-transgenic mice (n = 4).

- 38 -

Retrieving mouse APP exon 16 from genomic library

Phage lifts: The mouse 129 genomic library from Stratagene (cat#946308) was titred and plated out 20 150 mm LB plates containing ~50,000 phage/plate. Duplicate lifts were made from each plate using
5 Amersham Hybond-N+ nylon membranes. The plates were refrigerated for several hours to ensure the top agar was hardened. The membranes were placed atop the plaques and left on for 5 minutes. The membranes were lifted off the plates and placed plaque-side up on 3MM paper saturated with denaturation solution (0.1 M NaOH, 1.5 M NaCl) for 5 minutes. The
10 membranes were transferred briefly to dry 3MM paper to absorb the excess solution and then placed on 3MM paper saturated with neutralizing solution (0.2 M Tris-Cl pH 7.5, 2X SSC) for 5 minutes. The membranes were rinsed by placing them on 3MM paper saturated with 2X SSC for 5 minutes and then air dried. A digoxigenin-labeled mouse specific APP exon 16 probe of
15 93 bp was generated using PCR (from nt 1877 to 1969 in sequence MUSABPPA, accession #M18373).

PCR assay: In a 50 μ l total reaction volume was added 1 μ g genomic mouse tail DNA, 5 μ l 10X PCR buffer (Perkin Elmer cat#N808-0006), 5 μ l 2 mM dATP, dCTP, dGTP mix, 5 μ l 1.3 mM dTTP, 3.5 μ l 1 mM
20 digoxigenin-11-dUTP, 3 μ l 100 ng/ml oligonucleotide mix of KC65 (5'GTTCTGGGCTGACAAACATC3') and KC66 (5'GATGGCGGACTTCAAATCCTG3'), and 2.5 units AmpliTaq (Perkin Elmer cat#N808-0070). The reaction was run in a Perkin Elmer turbo 9600 thermal cycler. The parameters of the run were as follows: one cycle at
25 94°C for one minute, 36 cycles at 94°C for 30 seconds-56°C for 50 seconds-70°C for two minutes, maintain at 10°C indefinitely. Four individual PCR reactions were pooled and passed through a Sephadex G-50 column from Boehringer-Mannheim (cat#100616) in 10 mM Tris-Cl pH 7.5, 1 mM

- 39 -

EDTA, and 0.1 % SDS. Several dilutions of the dig-labeled probe were blotted onto a membrane and compared to standard amounts of a dig-labeled control DNA.

Hybridization of plaque-lifted membranes: Membranes were pre-
5 hybridized in 50% formamide, 5X SSC, 0.1 % N-lauryl sarcosine, 0.02 %
SDS, and 5% blocking reagent supplied by Boehringer-Mannheim
(cat#1096176) and incubated at 42°C for 4 hours. The pre-hybridization
solution was discarded and replaced with identical fresh hybridization solution
that contained 2 µg of the dig-labeled mouse APP exon 16 probe that was
10 boiled for 10 minutes and chilled on ice. Membranes were hybridized over a
two-day period at 42°C. All incubations (and heated-washes) were
performed in the Stovall "Belling Dancing" water bath. The
probe/hybridization solution was removed and saved for subsequent
screenings. Membranes were washed four times in 2X SSC, 0.1 % SDS at
15 room temperature for 5 minutes. Subsequent washings were as follows: two
washes of 30 minutes at 65°C in 0.5X SSC, 0.1 % SDS; two washes for 30
minutes at 65°C in 0.2X SSC, 0.1 % SDS; ten minutes at 65°C in 0.2X SSC;
and ten minutes at room temperature in 0.2X SSC.

Digoxigenin detection assay: The remaining protocol is taken from the
20 Boehringer-Mannheim "DIG Nucleic Acid Detection Kit" (cat#1175041).
Membranes were rinsed once for 2 minutes at room temperature in Genius 1
buffer (100 mM Tris-Cl, pH 7.5, 150 mM NaCl) and blocked for 1 hour at
room temperature in Genius 2 buffer (2% w/v blocking agent in Genius 1
buffer). Membranes were incubated with 150 µunits/ml of polyclonal sheep
25 anti-digoxigenin alkaline phosphatase conjugated antibody in Genius 2 buffer
for 30 minutes at room temperature. Two washes were done for 15 minutes
each at room temperature in Genius 1 buffer and once for 2 minutes in AP

- 40 -

9.5 buffer (100 mM Tris-Cl pH 9.5, 100 mM NaCl, 50 mM MgCl₂).

Membranes were processed in Lumi-Phos 530 (Boehringer-Mannheim cat#1275470) and placed in the dark for 16 hours then exposed to film for 20 minutes.

5 Positive plaques were picked and placed into 1 ml SM buffer (5.8 g NaCl, 2.0 g MgSO₄·7H₂O, 50 ml 1 M Tris-HCl pH 7.5 to a total volume of one liter) to diffuse and stored at 4°C. These plaques were screened in a PCR assay using the identical oligonucleotide pair that was used to generate the probe (assay-15 µl phage stock and 35 µl water were heated to 95°C for 10 20 minutes into which was added 10 µl 10X PCR buffer, 3 ml 100 ng/ml oligo mix of KC65 and KC66, 10 µl 2 mM dNTP mix, 5 units AmpliTaq, and 1 unit Perfect Match Polymerase Enhancer (Stratagene cat# 600129) to a total volume of 100 µl).

15 Secondary membrane screenings on 4 isolates were performed using the digoxigenin-mouse APP exon 16 probe previously made. Two positive phage plaques were grown (protocol taken from BioTechniques 7:21-23) to obtain enough DNA for further analysis.

20 A 15 Kb sequence containing the mouse APP exon 16 was sub-cloned into pBluescript IISK+ (Stratagene cat#212205) at the NotI site (designated as plasmid 35A) using standard cloning procedures. Southern analysis using a 32p-labeled mouse APP exon 16 probe revealed a 5 Kb NcoI fragment which became the backbone into which our human APP cDNAs were fused.

25 Southern analysis: Six separate reactions containing 1 µg of plasmid 35A were digested with 10 units each of restriction enzymes ApaI (cat#114S), ApaI/BglII (cat#144L), NcoI (cat#193L), NcoI/BglII, XbaI (cat#145S), XbaI/BglII (supplied by New England Biolabs) in their respective buffers (total volume of 30 µl) at their respective incubation temperatures for 3 hours. One-half of the digestion reactions was loaded onto an 0.8% agarose

- 41 -

(Bio-Rad cat#162-0133) gel in 1X TBE buffer. The gel was run at 20 volts overnight at room temperature. After photographing the gel, it was prepared for transferring to a nylon membrane. The gel was soaked in 0.25N HCl, rocked gently, for 15 minutes, rinsed well with water then soaked in 0.4N NaOH, rocked gently, for 20 minutes. A 3MM paper wick transfer was assembled using Amersham Hybond-N+ nylon membrane in 0.4N NaOH buffer overnight at room temperature. The membrane was rinsed in 5X SSC for 10 minutes at room temperature and UV cross-linked in a Stratalinker (Stratagene cat. #400071) using 1.2×10^5 mJoules for 30 seconds. The membrane was hybridized in 50% deionized formamide, 5X SSC, 0.1% N-lauryl sarcosine, 0.02% SDS, and 5% blocking agent (Boehringer-Mannheim) at 42°C, rocked gently, and incubated overnight. This solution was removed and replaced with the previously made mouse APP exon 16 digoxigenin-labeled probe (denatured) in fresh hybridization buffer and incubated at 42°C, rocked gently, for overnight. All subsequent washes, blocking, and antibody binding was identical to the protocol stated previously as digoxigenin detection assay.

Construction of the targeting vectors

Subcloning mouse exon 16 locus: The 5 Kb NcoI fragment containing the mouse APP exon 16 sequence was cloned into pBluescript IISK+ at an engineered NcoI site to generate pMTI2396 (Figure 2; see below). The 3 Kb 5'-flanking NcoI fragment and 2 Kb 3'-flanking NcoI fragments from p35A were also cloned into pBluescript IISK+ at the engineered NcoI site to generate pRA3 and pN2C4, respectively (Figures 3 and 4; see below). The pBluescript vector (1 μ g) was digested with 20 units of XbaI in buffer 2 (NEB) and incubated for 2 hours at 37°C. Ten units of calf intestine alkaline phosphatase (CIP from Boehringer-Mannheim cat#713023) were added to the reaction and incubated for 1 hour at 37°C to dephosphorylate the 5' ends. To

- 42 -

500 ng of the 5Kb NcoI fragment was added 6.2 pmol of phosphorylated, annealed adapter KC95/96 (5'CTAGACACTC3') using 400 units of T4 DNA Ligase (NEB cat#202L) in its appropriate buffer (50 mM Tris-Cl pH 7.8, 10 mM MgCl₂, 10 mM DTT, 1 mM ATP 25 mg/ml BSA) at 25°C for a 5-hour incubation. This reaction was digested with 20 units of XbaI and adjusted the buffer concentration to 50 mM NaCl and incubated for 1 hour at 37°C. The enzyme was heat inactivated at 65°C for 20 minutes. The DNA was removed from the residing enzymes using Strataclean resin (Stratagene cat#400714). To the 25 µl enzyme digestion reaction was added 5 µl of Strataclean resin, vortexed for 15 seconds and set at room temperature for 1 minute. It was then spun in an Eppendorf microcentrifuge 5415C at 14000xg for 1 minute. The supernatant was transferred to a clean tube and the procedure was repeated once. Dephosphorylated XbaI-linearized pBluescript, 50 ng, was combined with 500 ng of the phosphorylated 5Kb NcoI-adapter fragment in a standard ligation reaction and incubated at 14°C for overnight. The ligase was heat inactivated at 70°C for 10 minutes and one-tenth of the reaction was transformed into Epicurian.coli XL-1 blue cells (Stratagene cat#200236) using the protocol provided. The resulting construct having mouse genomic sequences for the 3' end of intron 15-exon 16-5' end of intron 16 was then referred to as p2396 (Figure 2). The BglII site within exon 16 is the point at which the human cDNA sequence was fused.

Subcloning of NcoI fragments proximal to the Exon 16 targeting site:

The 3 Kb intron 15 fragment and the 2 Kb intron 16 fragment were generated by a NcoI digestion on the template plasmid 35A (see Figure 1). The 3Kb and 2 Kb NcoI-NcoI fragments were then subcloned into the Bluescript (NcoI) vector (see above). The resulting plasmids were named pRA3 (3 Kb fragment; Figure 3) and pN2C4 (2Kb fragment; Figure 4). They were expanded and the 2 and 3 Kb fragments themselves isolated by GeneClean

- 43 -

(Bio 101). These isolated fragments were then used as probes in Southern blot paradigms.

Generation of cloning sites around the neomycin resistance gene: As an integral part of our targeting vector construct, we cloned the neomycin resistance gene (pPol2longneobpA provided by Ann Davis) downstream of our human APP cDNA sequence. The neomycin resistance gene (contained within a pBluescript KS+ vector) was under transcriptional regulation of the DNA polymerase II promoter sequence (long version) and the bovine growth hormone (BGH) polyadenylation sequences. Sequences composed of different restriction sites had to be cloned onto both the 5' and 3' ends of this gene construct. The plasmid, 2 μ g, was linearized with Sall, ligated to 45 pmol of annealed Sall-AflIII-EcoRV-NcoI-MluI adapter (5'TCGACGACTTAAGTTGATATCCACCATGGTGACGCGTT3') using 400 units of T4 DNA Ligase in its appropriate buffer at 14°C in an overnight incubation. This reaction was digested with EcoRV (cat#195S) and ligated to close. This plasmid, now referred to as p2395, was digested with XhoI to linearize it at the 3' end of the BGH sequence. Ligated to this XhoI site was an XhoI-BglIII-StuI adapter (5'TCGAGTGAGATCTTAAGGCCTGG3'). The ligase was removed from the reaction using the Wizard DNA clean up system (Promega cat#A7280) following the directions supplied in the kit. The linearized plasmid-adapter DNA (approx. 5 μ g) was digested with 30 units each, in one 50 μ l reaction, of StuI (cat#187L)/EcoRV in restriction enzyme buffer 2 (from NEB) at 37°C for 3 hours. The digest reaction was run through a 0.8% low melt agarose (FMC cat#50112) gel in 0.5X TAE buffer (20 mM Tris acetate, 0.5 mM EDTA) at 75V for 2 hours at room temperature. The 1800 bp band containing the promoter/neomycin/polyA sequences was excised from the gel and extracted from the agarose using the

- 44 -

Wizard DNA clean up system. This fragment was ligated to the human APP cDNA-adaptor generated through the follow process.

Generation of human APP cDNA's with either the Swedish-FAD, London-FAD, Swedish/London-FAD, or Swedish-FAD, APP713 mutations fused with human APP genomic sequences containing APP polyadenylation signals: Plasmid pMTI-2385-Swedish (not shown) possesses the entire human APP 695 cDNA fused with human APP cloned into pBluescript II SK+. The plasmid pMTI 2398 was derived from pMTI2385. The strategy for its creation involved the extensive use of a cDNA-genomic hybrid plasmid, pMTI2339. pMTI2385-Swedish was assembled in a four-part ligation with the following components; an ~1861 bps. XmaI-BglII fragment from pMTI2339, a ~2008 bps. SpeI-SalI fragment from pMTI2339, a ~589 bps. fragment from FAD clone #5 (contains Swedish-FAD mutation) generated by Dr. Gerhard Konig, and a pBSSK(+)II vector opened up with XmaI and SalI. The ligation was done according to standard protocols with the insert fragments being in equal molar ratios and there being a 3:1 ratio of total insert to vector. Ligation mixtures were transformed in XL-1 Blue competent cells (Stratagene) and mini-preps analyzed by an initial digestion of XmaI-SalI. Two putative clones were further characterized with BglII-SpeI, XmaI-BglII, SpeI-SalI, EcoRI, HincII, and PvuII. Two clones, #4 and #5 gave the expected results. These were grown up and sequenced confirmed.

Plasmid pMTI-2453 was derived from pMTI-2385-London. pMTI-2385-London was assembled in a four-part ligation with the following components: a ~1.7 Kb Xma I-SacI fragment from pMTI2385-Swedish, a ~350 bp SacI-StyI fragment from pMTI-104 (contains London-FAD mutation; obtained from Paul Fracasso), a ~2.5 Kb StyI-SalI fragment from pMTI2385-Swedish, and a ~2.7 kb SalI-XmaI fragment from pMTI2385-Swedish.

- 45 -

Plasmid pMTI-2454 was derived from pMTI-2385-Swedish/London. Swedish/London was assembled in a four-part ligation with the following components: a ~1.9 Kb XmaI-EcoRI fragment from pMTI2385-Swedish, a ~700 bp EcoRI-ClaI fragment from pMTI-2385-London, a ~1.9 Kb ClaI-Sall fragment from pMTI2385-Swedish, and a ~2.7 Kb Sall-XmaI fragment from pMTI2385-Swedish.

Plasmid pMTI-2455 was derived from pMTI-2385-Swedish APP713. pMTI-2385-Swedish APP713 was assembled in multi-step process using PCR mutagenesis to introduce the APP713 stop mutation into proximity with the Swedish-FAD mutation. First, a ~560 bp EcoRI-SpeI fragment from pMTI2385-Swedish was ligated with the 2.9 Kb EcoRI-SpeI fragment from Bluescript KS+II (Stratagene) to generate pMTI-X. A ~400 bp fragment containing the APP 713 stop mutation was generated by PCR using APP cDNA as template and oligonucleotides RA39 (CCATCGATGGATCAGTTACGGAAACGATGCTCTCATGC) and RA40 (CCATCGATGGCCAAGGTGATGACGATCACTGTGGATCCCTACGCTATGACAACACCGC) (Figure 9). The ~400 bp PCR fragment was digested with ClaI and StyI and ligated into the ~3.3 Kb ClaI-StyI fragment from pMTI-X to generate pMTI-Y. pMTI-2385-Swedish APP713 was assembled in a four-part ligation with the following components: a ~560 bp EcoRI-SpeI fragment from pMTI-Y, a ~1.9 Kb XmaI-EcoRI fragment from pMTI2385-Swedish, a ~2 Kb SpeI-Sall fragment from pMTI2339, and a ~2.8 Kb fragment from Bluescript SK+II.

Generation of the human APP "Swedish" FAD mutation cDNA-neomycin sequences to fuse to the mouse APP exon 16 DNA: Four µg of plasmid pMTI-2385B was digested with 20 units of restriction enzyme Sall (cat#138L) in its ideal buffer for 2 hours at 37°C. The reaction was run through an 0.8% agarose gel in 0.5X TAE buffer at 120 volts for 1.5 hours at

- 46 -

room temperature. The linearized DNA band was excised and isolated away from the agarose using the Qiaex DNA Gel Extraction kit and the protocol provided (Qiagen Cat#20021). One μg of SalI-linearized p2385B was ligated to 45 pmol of annealed SalI-AflIII-EcoRV-NcoI-MluI adaptor (mentioned previously) in a standard ligation reaction. One-tenth of the ligation reaction was used to transform E.coli XL-1 blue cells. This p2385B-adaptor construct, 18 μg , was linearized with 60 units of EcoRV in a standard digestion reaction. Into this was added 15 units of calf intestine alkaline phosphatase and incubated at 37°C for 1 hour to dephosphorylate the 5' ends of the DNA. The reaction was stopped with EDTA at a final concentration of 5 mM and heat inactivated at 75°C for 10 minutes. The dephosphorylated plasmid was gel isolated and 1 μg was ligated to 400 ng of the 1800 bp neomycin fragment with EcoRV 5' and StuI 3' ends (mentioned in the section "Generation of cloning sites around the neomycin resistance gene"). One-tenth of the ligation reaction was used to transform E. coli XL-1 blue cells following the protocol provided by the supplier. Correct orientation constructs had the neomycin fragment (5' EcoRV site) placed immediately downstream of the human APP cDNA polyA sequences (3' EcoRV site), this construct was designated p2397 + A (not shown).

Construction of the completed targeting vector containing the human APP "Swedish" FAD mutation: The 5 Kb mouse APP exon 16 containing DNA, p2396 (12 μg), was digested with 50 units of BglII in buffer 3 for 3 hours at 37°C. To 6 μg of the digest was added 10 units of CIP and incubated at 37°C for 1 hour. The reaction was stopped as mentioned above and the DNA was gel isolated using Gelase (Epicentre cat#G09100) and following the supplied protocol. The 4.5 Kb BglII fragment containing the human APP cDNA-neomycin fusion was released from p2397 + A by digesting 12 μg of DNA with 28 units of NruI (cat#192L) in its ideal buffer at

- 47 -

37°C for 3 hours. After being confirmed of its linearization 50 units of BglII were added for an additional 2-hour incubation at 37°C. The 4.5 Kb fragment was gel isolated using Gelase and then ligated (300 ng) to the dephosphorylated p2396-BglII linearized DNA (100 ng) in a standard ligation reaction and subsequently transformed into E.coli XL-1 blue cells. The resulting plasmid with the mouse APP exon 16 fused to the human APP cDNA at exon 16 (BglII site) was designated as p2398 (Figure 5 and Figure 18).

E.2.7 Cloning of the HSV thimidine kinase (TK) gene into the targeting vector: The HSV thimidine kinase gene (from pAD7) was provided by Ann Davis. Unique restriction sites had to be engineered with the TK gene to provide linearizing access in the completed targeting vector. A 3Kb BamHI-ClaI fragment containing the murine phosphoglycerate kinase (PGK) promoter regulating the TK gene with the BGH polyadenylation sequences was isolated away from vector sequences and sub-cloned into pBluescript II SK+ at its respective sites. Twenty µg of this new TK plasmid, pCB11, was digested with 60 units of Sall in its unique buffer and incubated overnight at 37°C. The enzyme was heat inactivated at 65°C for 20 minutes and then 10 units of CIP was added for 1 hour at 37°C. The phosphatase was heat inactivated at 75°C for 10 minutes. The linearized DNA band was excised and isolated away from the agarose using the Qiaex DNA Gel Extraction kit and the protocol provided as stated above. In a standard ligation reaction, 45 ng of Sall-linearized vector was added to 15 pmol of annealed Sall-AscI-PmeI-NotI-AscI-PmeI-Sall adapter (5'TCGACAAGGCGCGCCGTTTAAACAAGCGGCCGCTTGGCGCGCCTTTTGTTTAAACTTG3') and incubated overnight at 14°C. This TK plasmid containing the restriction sites PmeI and AscI was designated as pXII28N. Five µg of pXII28N was digested with 20 units of NotI (cat#189L) and 15 units of PvuI (cat#150L) in NotI buffer (NEB) and 0.1 mg/ml BSA at 37°C

- 48 -

for overnight. The 3Kb NotI TK band was excised and isolated away from the agarose using the Qiaex DNA Gel Extraction kit and the protocol provided. Five hundred ng of this TK fragment was ligated to 50 ng of NotI linearized p2398 (vector containing the APP/neo sequences fused to the mouse APP exon 16 sequences) in a standard ligation reaction and incubated overnight at 14°C. The resulting targeting vector, p2399 (350 µg), was linearized with 320 units of PmeI (cat#560L) in buffer 4 (NEB) and 0.1 mg/ml BSA and incubated overnight at 37°C. Protein was removed by adding sodium acetate pH 5.2 to 0.3 M and extracting twice with Tris-Cl buffered phenol and extracting once with chloroform and ethanol precipitating at -20°C for overnight.

Construction of the completed targeting vectors containing the human APP London-FAD, Swedish/London-FAD, and "Swedish-FAD APP713

mutation: These three targeting vectors were constructed by ligating four separate fragments with one of these fragments containing one of the FAD mutations. The seminal targeting vector construct, p2398, was digested in three independent reactions to obtain three of the specific fragments.

(1.) Five µg of p2398 were digested with 15 units of AflII (cat #520S) in buffer 2 with 0.1 mg/ml BSA at 37°C for overnight. To this reaction was added enough buffer 3 to adjust the concentration to 100 mM NaCl and 30 units of NotI and incubated at 37°C for 3 hours. (2.) Twenty µg of p2398 were digested with 24 units of BglII, 20 units of NotI, and 0.1 mg/ml BSA in buffer 3 at 37°C for overnight. (3.) Twenty µg of p2398 were digested with 20 units of AflII, 10 units of ClaI (cat#197L), and 0.1 mg/ml BSA in buffer 4 at 37°C for overnight. All three digestion reactions were run on 0.8% low melt agarose gels in 0.5X TAE buffer at 70 volts for 3 hours. From digestion reaction (1.) an 8Kb fragment containing the neomycin-murine APP intron 16-pBluescript sequences was excised, from reaction (2.) a 2Kb

- 49 -

fragment containing murine APP intron 15-exon 16 sequences was excised, from reaction (3.) a 2Kb fragment containing human cDNA/polyA sequences was excised and all the fragments were isolated away from the agarose using the Qiaex Gel Extraction kit. The last fragments to isolate were the three

5 700bp human APP FAD containing fragments. Twenty-five μ g of each of the pMTI-2385 London (not shown), pMTI-2385 Swedish/London (not shown), and pMTI-2385 Swedish-FAD 713 (not shown), vectors were digested with 24 units of BglII, 15 units of ClaI, and 0.1 mg/ml BSA in buffer 4 at 37°C for overnight. The 700 bp bands from these digestions were isolated away

10 from the agarose using the identical protocol as above. A four-part standard ligation reaction was combined using 25 ng of the 8 Kb AflII/NotI fragment, 250 ng of the 2 Kb AflII/ClaI fragment, 300 ng of the 700 bp BglII/ClaI fragment, and 250 ng of the 2 Kb NotI/BglII fragment and incubated at 14°C for 24 hours. One-sixth of the ligation reaction was used to transform E.coli

15 XL-1 blue cells in a standard protocol. The resulting constructs were designated as p2450 (London-FAD), p2451 (Swedish/London-FAD), and p2452 (Swedish-FAD APP713)(not shown). The final step for each individual plasmid was to clone the TK gene fragment with NotI ends into it. Five μ g of each plasmid, p2450, p2451, p2452 were digested with 20 units of

20 NotI in buffer 3 at 37°C for 3 hours. To dephosphorylate the vector, 10 units of CIP were added to the digestion reaction and incubated at 37°C for 1 hour. The phosphatase was heat inactivated at 75°C for 10 minutes. The linearized DNA band was excised and isolated away from the agarose using the Qiaex DNA Gel Extraction kit and the protocol provided as stated above.

25 Fifty ng of each dephosphorylated vector was ligated to 300 ng of the 3 Kb NotI TK gene fragment in a standard ligation reaction. The resulting plasmids were designated as p2453 (London-FAD; Figure 6: Figure 19), p2454 (Swedish/London-FAD; Figure 7; Figure 20), and p2455 (Swedish-FAD APP713; Figure 8: Figure 21). Each of these three targeting vectors

- 50 -

(500 μ g) were linearized with 500 units of AscI in buffer 4 at 37°C for overnight. The DNAs were cleaned away from the enzymes by phenol/chloroform extractions as stated in the section "Cloning of the HSV thimidine kinase (TK) gene into the targeting vector". Linearized plasmids were electroporated into ES cells.

miniSouthern-blot analyses

DNA sample preparation: Potential clones were grown in a 96 well plate format. Samples were lysed with the addition of 50 μ l of Lysis Buffer [10 mM Tris pH 7.5, 10 mM EDTA pH 8.0, 10 mM NaCl, 0.5% Sarcosyl, and 1 mg/ml Proteinase K (added fresh)] per well and incubated overnight at 65°C in a humidified chamber. The DNA is precipitated by the addition of 100 μ l of 75 mM NaCl in ethanol followed by incubation at room temperature for 15-30 minutes. The DNA is then washed 3x with 150 μ l of 70% ethanol added drop by drop to each well. After the final wash, the plate is inverted and allowed to air-dry for 5 - 10 minutes. While the plate is drying, the Restriction Enzyme Cocktail (1x Restriction Buffer specified for the enzyme being used, 1 mM Spermidine, 100 μ g/ml Bovine Serum Albumin, and 10 - 20 units of enzyme) is prepared. 30 μ l of this cocktail is then added to each well. Incubate overnight at the restriction enzyme's required temperature in a humidified chamber. The next day add 4-5 ml of loading dye (10 mM Tris-HCl pH 6.0, 0.25% bromophenol blue, 0.25% xylene cyanol FF, 15% Ficoll (Type 400; Pharmacia) in water, and 30 mM EDTA.) and store at -20°C.

Agarose Gel Electrophoresis: A large gel tray (Owl Scientific) is prepared with three 36-teeth combs (evenly distributed along the length of the tray) and 400 ml of molten agarose (FMC). This size of gel will accommodate one 96-well mini-Southern digest plate. The samples were

- 51 -

electrophoresed for approximately three hours at 120 V. After the electrophoresis was complete, the gel was denatured in 0.25 M HCl (2 x 7 minutes at room temperature) and then equilibrated in 0.4 N NaOH (1 x 20 minutes at room temperature). An overnight alkaline capillary transfer is set up in 0.4 N NaOH with Gene Screen Plus (DuPont NEN). The next day the membrane was neutralized in 2x SSC for 5-10 minutes and then UV cross-linked (Stratagene). The membrane is then stored dry until hybridization. Prehybridization was carried out in 1 M NaCl (Gibco BRL), 10% Dextran Sulphate (Pharmacia), 1% SDS (Gibco BRL), and 200 µg/ml salmon sperm DNA(Stratagene) for at least one hour at 65°C in a Robbins Hybridization Oven. The probe of interest was then labeled according to the standard protocol contained in the Prime-It II random prime kit (Stratagene). The specific activity of the probe was approximately 1 x 10⁹ dpm/µg. It was then added directly to the prehybridization mixture at a concentration ~1 x 10⁶ dpm/ml. The filter(s) were then hybridized for 16 hours at 65°C in the hybridization oven. The initial post - hybridization wash was carried out for 5-10 minutes at room temperature in 2x SSC (3 M NaCl, 0.3 M Sodium Citrate Dihydrate), 1% SDS. A stringent wash was then performed in 1x SSC, 0.1% SDS at 65°C for 30 minutes. The filter was then placed into a seal-a-meal bag and placed into a Fuji Phosphoimager for interpretation.

Confirmatory Southern Blots

Preparation of High Molecular Weight DNA from Cells: In order to confirm targeted clones identified in the mini - Southern paradigm, cell pellets expanded from these clones are analyzed for accurate recombination events at both the 5' and 3' ends of the targeting vector. 1 ml of Cell Lysis Buffer (100 mM NaCl, 50 mM Tris pH 7.5, 10 mM EDTA pH 8.0, and 0.5% SDS) and 20 ml of freshly prepared 40 mg/ml Proteinase K (Boehringer-Mannheim) was added to each cell pellet. The tubes were

- 52 -

rocked overnight at 65°C. The next day, an equal volume amount of isopropanol was added and the tube inverted several times to precipitate the DNA. The DNA was then spooled onto a flame sealed micropipette and rinsed once in 70% ethanol, once in 100% ethanol, and then air dried. The pipette was broken off into a sterile Eppendorf tube and the DNA dissolved in 200 µl of sterile TE overnight at room temperature. The DNA is then stored at 4°C until restriction enzyme analysis.

Agarose Gel Electrophoresis: Restriction enzyme digested DNA is gel analyzed as described above in the mini-Southern methods except the number and sizes of the combs vary. Denaturation, renaturation, and capillary transfer were performed as described previously. Probes of interest were also labeled in the same manner as described above. Interpretation of results were facilitated by phosphoimaging as previously described.

Gene-targeting in ES cells

Culture of ES cells: Procedures were performed essentially as describe in E. J. Roberstion (Robertson, 1987) . ES cell were propagated using Mitomycin C treated SNL76/7 STO feeder cells (cell line obtained from A. Bradley) and modified DMEM culture media (supplemented with 15% FCS, 1X GPS, 1X BME).

Electroporation of ES cells: DNA was linearized with the appropriate restriction enzyme then extracted with an equal volume of phenol/chloroform and once with an equal volume of chloroform and precipitated with 2.4 volumes of ethanol. The DNA was resuspended at 1 mg/ml in sterile 0.1X TE (25 ml of DNA per electroporation). Embryonic stem cells (80% confluent) were passaged 1:2 the day before electroporation. Cells to be electroporated were fed 4 hours before harvesting. The cells were trypsinized and resuspend in media (cells from 2 x 10 cm plates can be combined in a total volume of 10 ml in a 15 ml tube). The cells were pelleted and resuspend in 10 ml PBS at a density of 11×10^6 cells/ml. The appropriate amounts of DNA and cells were mixed together in a 15 ml tube (25 ml of DNA and 0.9 ml of cells for each electroporation) and allowed to sit at room temperature for 5 minutes. The cell/DNA mixture (0.9 ml) was transferred to electroporation cuvettes and an electrical current was passed through the solution (using Biorad GenePulser at 230V and 500 mF). The cells were then transferred to culture plates with feeder cells (up to 2×10^7 cells/100 mm plate or 6×10^6 cells/60 mm plate). After 24 hours of culture in modified DMEM the cells were cultured in DMEM selection containing G418 and 0.2 mM FIAU. Resistant colonies may be picked as early as 8 days, are best around 10-11 days, but may be recovered up to 18-21 days after the electroporation. Picked colonies are transferred to 96 well plates with feeders

- 54 -

cells and screened for gene-targeting events by mini-Southern-blot analysis (see below).

Production of chimeric mice: Procedures were performed essentially as described by A. Bradley (Bradley, 1987). Host 3.5 day blastocysts were derived from timed matings of C57BL/6 mice and cultured in M16 media. Approximately 14 targeted ES cells were injected into each blastomere. Surviving blastocysts were then surgically reimplanted (approximately 12 per animal) into pseudopregnant ICR female mice essentially as described (A. Bradley). Chimeric mice were born about 17 days after implantation.

Genotype analyses of transgenic mice

Identification of mice possessing the targeted human APP cDNA by PCR screening: When mice were older than 2 weeks of age their tails were biopsies to obtain genomic DNA for analysis. One centimeter pieces of tail were prepared using the QIAamp Tissue Kit (Qiagen cat# 29304) and following the protocol provided. Genomic DNA was eluted in 150 μ l of 10 mM Tris-Cl pH 9 and used in two independent PCR assays; (1) to determine the endogenous mouse APP allele that remained intact: total reaction volume of 50 μ l - 5 μ l of genomic tail DNA (approximately 1 μ g), 5 μ l of 10X buffer 8 (Stratagene cat#200430), 5 μ l of 2 mM dNTP mix, 200 ng of oligonucleotide KC125 (5'ACTTTGTGTTTGACGC3'), 200 ng of oligonucleotide KC132 (5'CAGTTTTTGGATGGCGG3'), 1 unit of Perfect Match Polymerase Enhancer, 2.5 units of AmpliTaq and 100 ng each of oligonucleotides 6&7 and (2) to determine the targeted mouse APP allele: total reaction volume of 50 μ l - 5 μ l of genomic tail DNA (approximately 1 μ g), 5 μ l of 10X buffer 8 (Stratagene cat#200430), 5 μ l of 2 mM dNTP mix, 200 ng of oligonucleotide KC125 (5'ACTTTGTGTTTGACGC3'), 200 ng of oligonucleotide KC131 (5'GATGATGAACCTTCATATCCTG3'), 1 unit of

- 55 -

Perfect Match Polymerase Enhancer, 2.5 units of AmpliTaq and 100 ng each of oligonucleotides 6&7. The reactions were run in a Perkin Elmer turbo 2400 thermal cycler. The parameters of the run were as follows: one cycle at 94°C for one minute, 30 cycles at 94°C for 30 seconds-56°C for 50 seconds-70°C for two minutes, maintain at 10°C indefinitely.

Oligonucleotides 6

(5'CCTCGGCCTTTGGTGTGTGTTTTATGACATGACCCCCTTGA) & 7

(5'CACCCTGTTGTCAATGCCTCTGGGTTTCCGCCAGTTTCG3') are

homologous to mouse ribosomal protein L32 sequences within intron 2 and exon 3, respectively, and used as an internal DNA control signal. One-fifth of each PCR reaction was run on a 6% polyacrylamide gel (Novex cat#EC6265) in 1X TBE (89 mM Tris borate, 2 mM EDTA) buffer at 125 volts for 35 minutes and stained in 1 mg/ml EtBr for 15 minutes and photographed.

RNA analyses

RNA isolation: Total brains were dissected and flash frozen on dry ice from two negative litter mates, two heterozygous targeted mice, and two homozygous targeted mice. In addition, kidneys and tails were also removed from these mice and flash frozen. The brains were divided in half, one for the RNA analysis and the other for protein analysis. To one-half of each brain was added 5 ml RNAzolB (Tel-Test, Inc. cat#CS-105) and the tissues were homogenized using a Brinkman Polytron at medium speed for 20 seconds. Chloroform, 500 µl, was added to the homogenized tissue and shaken well for 10 seconds and incubated on ice for 15 minutes. The samples were spun in a tabletop Sorvall centrifuge at 1500Xg for 20 minutes at 4°C. The aqueous phase was removed and added to an equal volume of isopropanol, mixed, and incubated on ice for 15 minutes. The samples were spun in a Sorvall RC-5B centrifuge with an SS-34 rotor at 7500Xg at 4°C for

- 56 -

25 minutes. The supernatants were removed and the pellets were rinsed twice in cold 70% EtOH and air dried. The total RNA pellets were resuspended in 500 μ l H₂O and incubated at 65°C for 10 minutes to more easily get the RNA into suspension. These RNA samples were used to obtain polyadenylation specific mRNA using the PolyAtract mRNA isolation system III kit (Promega Z5300). The protocol followed was provided by the supplier and yields ranged from 3 to 6 μ g of mRNA.

Northern blot analyses: These samples were then used in a Northern blot to see the sizes of these targeted hybrid APP transcripts. The RNA was run on a 1.2% agarose (FMC cat# 50072), 2.2 M formaldehyde gel prepared as follows: 0.6g agarose in 36 ml H₂O were melted in a microwave and placed at 60°C. When the gel cooled to 60°C, 5 ml of 10X MOPS (0.4 M MOPS (Sigma MESA M-5755) pH7, 0.1 M sodium acetate, 10 mM EDTA pH8) running buffer and 9 ml of 37% formaldehyde (pH > 4) were added, mixed and left at 45°C until ready to pour. The RNA samples were prepared in a total volume of 30 μ l - 3 μ l 10X MOPS buffer, 5.25 μ l 37% formaldehyde, 15 μ l formamide, and 6.75 μ l mRNA (0.5 μ g) were mixed well and incubate at 55°C for 15 minutes. To this was added 6 ml formaldehyde loading buffer (1 mM EDTA pH8, 0.25% bromophenol blue, 0.25% xylene cyanol, 50% glycerol) and 1 ml 1 mg/ml EtBr. The samples were loaded into the gel and run at 5V/cm (55-75V) for 3hr in 1X MOPS buffer. The gel was rinsed in H₂O several times and soaked in 0.05N NaOH for 30 minutes under gentle shaking. The gel was then equilibrated twice for 15 minutes in 20X SSC and transferred by wick assembly for 16 hours in 20X SSC. The membrane used for transferring was Hybond-N+ (Amersham cat#RPN2020B) which is a 0.45 micron nylon membrane. After transfer was completed the membrane was rinsed in 2X SSC for 10 minutes and UV cross-linked in a Stratalinker mentioned earlier. The membrane was pre-hybridized

- 57 -

in 10 ml 0.5 M sodium phosphate pH7, 1 % BSA, and 7 % SDS for 4 hours at 65°C in a Robbins Hybridization Oven (model 400). This solution was removed and replaced with fresh hybridization solution and 6x10⁷ counts of denatured APP probe and 8x10⁵ counts of denatured mouse beta-actin probe and hybridized overnight at 65°C. The membrane was washed in 2X SSPE, 0.1 % SDS at 25°C for 10 minutes, twice, and then washed in 1X SSPE, 0.1 % SDS (pre-warmed) at 65°C for 15 minutes. The membrane was exposed to a phosphoimaging screen for 24 hours and developed.

Probes for Northern blot: Both the APP probe (homologous to the murine and human sequences) and the murine beta-actin probe were prepared in identical protocols. The APP DNA used to make the probe was an NruI/XhoI 900bp fragment from p2385B. The murine beta-actin 430bp DNA used for the probe came from a PCR reaction where the exon 3 of B-actin was amplified using these two oligonucleotides: KC137 (5'GTTTGAGACCTTCAACACCC3') and KC138 (5'GAAGGAAGGCTGGAAAAGAGCC3'). The probes were labeled using the Prime It II kit (Stratagene cat#300385) and following the protocol provided. After the reactions were stopped they were put over a G-50 spin column (5'-3' cat# 5303-633329) to remove the un-incorporated nucleotides.

To increase the level of APP-specific mRNA from the polyA selected RNA, the samples were annealed to an APP specific oligonucleotide (RA49-5'CGATGGGTTAGTGAAGCA3')) that was homologous to both the murine and human sequences approximately 40nt 3' of the stop codon. The assay was performed using the Superscript II RT-PCR kit (Gibco/BRL cat# 18089-011). In a reaction volume of 14 µl was combined 0.1-0.15 µg polyA mRNA and 600 ng RA49 and incubated at 70°C for 10 minutes and 4°C for 10 minutes. To this was added 2 µl 10X synthesis buffer, 1 µl 10X dNTP mix, 2 µl 0.1 M DTT, and 200 units of Superscript II reverse transcriptase (all

- 58 -

supplied by the kit) and the incubations continued at 25°C for 10 minutes, 42°C for 50 minutes, 70°C for 15 minutes, and 4°C for 10 minutes. The reactions were then treated with 2 units of RNaseH for 20 minutes at 37°C and then placed on ice. After the RNA was removed from the cDNA the

5 next step was the amplification reaction: 20 μ l of cDNA reaction mix, 8 μ l 10X synthesis buffer, 300 ng of oligonucleotide KC56 (5'GTGAAGATGGATGCAGAATTC3'), 300 ng of oligonucleotide KC56 Swedish (5'GTGAATCTAGATGCAGAATTC3'), 600 ng of oligonucleotide RA49, and 5 units of AmpliTaq in a total volume of 100 μ l. The

10 amplification was run in the Perkin Elmer turbo 2400 using the same parameters as stated in "Identification of mice possessing the targeted human APP cDNA by PCR screening". The RT-PCR reactions were subjected to restriction enzyme digestions taking advantage of the restriction site polymorphism between the murine and human APP sequences. One-tenth of

15 the RT-PCR reaction was digested with 30 units of SalI and 0.1 mg/ml BSA in its ideal buffer at 37°C for 2 hours, another set was digested with 30 units of StyI in buffer 3 at 37°C for 2 hours. The digests were run out on a 4% polyacrylamide gel in 1X TBE at 150 volts for 1 hour and stained in 1 mg/ml EtBr for 15 minutes and photographed. All oligos were provided by

20 Midland and all restriction enzymes by NEB.

Protein Analysis

Tissue Extraction: This protocol is generally used for mouse tissue with no more than several hundred mgs of tissue available, therefore all volumes must be kept to a minimum. Tissue was homogenized in 1 ml of

25 RAB buffer (0.1 M MES pH 7.0, 0.75 M NaCl, 0.5 M MgCl₂, 1 mM EGTA, 1 mM DTT) containing proteinase inhibitors. The protease inhibitor cocktail contains 1x Aprotinin (0.41 trypsin inhibitor units/mg protein), 1x PMSF (2 mM in isopropanol), 1x Protease inhibitor mix (chymostatin,

- 59 -

leupeptin, antipain, and pepstatin) each at 50 μ g/ml in DMSO, and 1 mM EDTA. A 7 ml dounce tissue grinder (Wheaton) was used for homogenization. The tissue homogenate was spun at 40K in the Beckman TL100 using the fixed angle rotor for one hour. The supernatant from this spin was saved as it contains the soluble APP. The pellet was homogenized in 1 ml of RAB plus protease inhibitors and 30% sucrose (Sigma). Spin for one hour at 40K in the Beckman TL100. This serves as a wash and demylelinating step. Discard the supernatant from this spin and homogenize the pellet in 1 ml of RIPA buffer (150 mM NaCl, 1 % NP-40, 0.5 % deoxycholate (Na salt), 0.1 % SDS, and 50 mM Tris-Cl pH 8.0). This should contain the membrane associated form of APP. The amount of protein can then be quantitated by using the BCA Protein Assay Reagent Kit (Pierce). This quantitation allows equal amounts of total protein to be loaded on polyacrylamide gels and direct comparisons of transgenic and non-transgenic expression patterns and levels to be made.

Immunoprecipitation: The final adjusted volume of the immunoprecipitation was 1 ml in RIPA buffer. The amounts of antigen and antibody to add varied from experiment to experiment depending on the concentrations of both. Antibody and antigen were incubated for two hours at 4°C while gently spinning on a rotating wheel. 50 μ l of goat anti-mouse or anti-rabbit IgG bound to agarose (Sigma) was added to the antigen/antibody and incubated for another two hours at 4°C on the rotating wheel. Agarose IgG-antigen/antibody complex was rinsed by pelleting at 12,000 x g for 1 min. and then removing the supernatant. Then 500 μ l of ice cold RIPA buffer was added to the pellet, resuspended, and incubated for 10 minutes on ice. The samples were then spun at 4°C. The rinses were repeated twice more, but the 10 minute incubation step was omitted. To the rinsed pellet, was added 50 μ l of 1x sample buffer (Novex) plus 2 ml of beta-

- 60 -

mercaptoethanol (Aldrich). Samples were boiled for 10 minutes and spun for 1 minute at room temperature. The supernatant was transferred to a fresh tube and store at -20°C.

Western Blotting: Polyacrylamide gel electrophoresis (PAGE) and electroblotting were accomplished utilizing the X-Cell II Gel and Blot Module (Novex) and pre-casted polyacrylamide gels (Novex). The selection of a particular separation scheme depended on what form of the Alzheimer Precursor Protein (APP) was being examined. For C-terminal fragments 16% Tris-Tricine gels were utilized, holo APP utilized 10-20% Tris-Tricine gels, and to elucidate form differences (Kunitz vs. 695) of the holo-APP, 6% Tris-Glycine gels were used. Samples prepared as described above were loaded onto gels and electrophoresed at 120V for approximately 90 minutes. The gels were then transferred to nitrocellulose membranes (Novex) for 1-2 hours at 30V. Non-specific sites were then blocked by incubation of the filter in 5% non-fat dry milk (NFDM) for 1 hour at room temperature while gently rocking. Primary antibody was then added at a dilution of 1:500 in 5-10 ml of NFDM, added to the membrane and sealed in a seal-a-meal bag. This was incubated overnight at room temperature while gently rocking. The membrane was then rinsed for 1 hour at room temperature with several changes of 5% NFDM. A ³⁵S labeled secondary antibody (Amersham), either anti-mouse IgG or anti-goat IgG, was then added and incubated for 1 hour at room temperature while gently shaking. The membrane was then rinsed for 15-30 minutes in 5% NFDM and then equilibrated into 1x phosphate buffered saline (PBS, Gibco BRL) for 15 minutes. The filter was then dried and either placed on a phosphoimaging plate or with a piece of X-OMAT X-ray film (Kodak).

- 61 -

APP Antibodies: Monoclonal antibody (MAb) 4G8 (Senetek) was used for the immunoprecipitation of APP holoprotein and C-terminal fragments at a dilution of 1:100 (~10-20 µg/ml). MAb 286.8 (BRC) was used for the immunoprecipitation of APP holoprotein at a dilution of 1:100 (~10-20 µg/ml). MAb 6E10 (Senetek) was used as a detection reagent on Western blots at dilutions of 1:500. Polyclonal antibody (PAb) 369 (generously provided by Dr. Sam Gandy) was used for both the immunoprecipitation of APP holo-protein (1:100) and for a detection reagent for C-terminal fragments (1:500). MAb 22C11 (generously provided by Dr. Konrad Beyreuther) was used as a detection agent for APP holo-protein at a dilution of 1:500.

FAD-m/hAPP gene products expressed in transgenic mouse lines

Transgenic mouse lines ES5007, ES5103, ES5401, and ES5403 were generated by mutating the mouse APP gene via homologous recombination in embryonic stem (ES) cells (see below). The gene products expressed in the transgenic mouse lines are described schematically in Figure 10. m/hAPP770 represents the largest (770 amino acid residues) of the various alternative splice forms of protein expressed by each mutated mouse APP gene. m/hAPP exhibits amino acid sequence identity with mouse APP with the exception of those residues indicated by (asterisks, *). In all cases the beta-amyloid (bA4) domain (Asp672 to Thr714; 43 amino acid residues) has been "humanized" by the introduction of three amino acid substitutions (as indicated by green asterisks); Gly(676) to Arg, Phe(681) to Thr, and Arg(684) to His. Transgenic mouse line ES5007 also has the Swedish-FAD mutation [Lys, Met(670, 671) to Asn, Leu] introduced into the mouse gene. Transgenic mouse lines ES5401 and ES5403 have the London-FAD mutation [Val(717) to Ilu] and transgenic line ES5103 carries both Swedish and London FAD mutations. In addition to the Swedish FAD and "human"

- 62 -

mutations, a transgenic mouse line ES5215 can also be produced which has a premature stop codon introduced at position 714.

Gene-Targeting Vectors

The targeting vectors were designed in such a way as to facilitate the integration of human cDNA sequences into mouse exon 16. The targeting constructs function as replacement-type vectors with both positive (neomycin resistance gene) and negative (HSV TK gene) selection genes (figure 11). To facilitate homologous recombination, a mouse genomic clone encompassing exon 16 was obtained by screening a mouse genomic lambda library. A lambda clone, "35A", was identified which contained an intact exon 16 (figure 1). Nco I fragments of mouse genomic clone 35A were subcloned into the BSII SK+ vector and the subclones (pRA3, pMTI-2396, and pN2C4) were characterized by DNA sequence and restriction enzyme analyses (see Figures 2, 3, and 4). The 5.5 Kb Nco I DNA fragment (subcloned into pMTI-2396) contains APP exon 16 and ~1.9 Kb and ~3.5 Kb from introns 15 and 16 respectively (Figure 2). The Nco I DNA fragment, containing exon 16, was the template upon which the gene-targeting vectors were constructed.

The gene-targeting vectors were designed so that mouse exon 16 gene sequences were fused (at a common Bgl II site) with human cDNA sequences which encode the remainder of exon 16 and exons 17 and 18 (figure 11). The mouse and human cDNA sequences encode the identical protein sequence with the exception of 3 amino acid differences (shown as green asterisks) which reside within the beta-amyloid domain. The mouse genomic-human cDNA fusion effectively "humanized" the beta-amyloid domain and facilitated the introduction of specific FAD mutations while leaving the remainder of mouse APP protein sequences unchanged (see Fig 13).

- 63 -

The human cDNA was mutagenized to encode either the "Swedish"-FAD, "London"-FAD, "Swedish"/"London"-FAD (shown here), or "Swedish"-FAD APP713 mutations (shown as black asterisks) of APP (see Fig. 10 and Table I). The mutagenesis of the "Swedish"-FAD mutation also incorporated a new Xba I restriction site. Proper RNA processing was ensured by fusing the 3'-end of the human cDNA sequence with human genomic sequences which contain transcription termination and polyadenylation signals from the human APP gene. A neomycin gene was inserted in-between the 3'-end of the human APP polyadenylation signal and mouse APP intron 16 sequences. Targeting vector pMTI-2398 (Swedish-FAD) contained the neomycin resistance gene and not the HSV TK gene. This vector was linearized with Pme I and was used to generate transgenic mouse line ES5007. (Tables I and II).

For the remaining three targeting vectors, a HSV Tk gene was inserted into the clone in such a way that its placement was outside of the genomic domains homologous to mouse (Figure 11; as shown or in the opposite orientation; the orientation was not critical). Targeting vector pMTI-5453 encodes London-FAD m/hAPP, targeting vector pMTI5454 encodes Swedish/London-FAD m/hAPP, and targeting vector pMTI-5455 encodes Swedish-FAD m/hAPP713. These targeting vectors were linearized with Asc I and were used to generate transgenic lines. (Tables I and II).

Gene-targeting in embryonic stem (ES) cells

The targeting vectors were designed to function as replacement-type vectors with both positive (neomycin resistance gene) and negative (HSV TK gene) selection genes. After electroporation of the targeting vector into ES cells, G418 drug treatment selected for ES cells which had integrated the targeting vector (including the neomycin resistance gene) into the mouse genome. The majority of G418 resistant ES cell clones had targeting vector

- 64 -

integrated at random locations of the genome. These ES cell clones retained an intact HSV TK gene and were not desired. The clones containing random integrations could be eliminated by treatment with FIAU selection media which is toxic only to cells expressing HSV TK. If, as desired, the mouse APP gene is targeted via a double-crossover homologous recombination event, the flanking non-homologous HSV TK DNA sequences are lost (as shown in fig 12) and the ES cells are resistant to FIAU treatment.

Homologous recombination between mouse APP exon 16 locus and the gene-targeting vector fundamentally alters the manner by which the gene encodes APP (see figure 13). Normally, the beta-amyloid, transmembrane, and cytoplasmic domains of mouse APP are encoded by three separate exons. In addition, the coding region for the beta-amyloid domain resides both on exons 16 and 17. After homologous recombination with the gene targeting vector, however, mouse exon 16 gene sequences are fused with human cDNA sequences. Mouse exons 17 and 18 are now displaced down-stream from the neomycin resistance gene and are inactive. The human cDNA now functions in place of mouse exons 16, 17, and 18 to encode APP. Therefore, the beta-amyloid, transmembrane, and cytoplasmic domains of mouse APP are now encoded by human cDNA sequences.

The gene products of this new mouse genomic-human cDNA fusion are designated m/hAPP. Human cDNA sequences (exons 16, 17, and 18) encode the identical protein sequence with the exception of 3 amino acid differences (shown as green asterisks) which reside within the beta-amyloid domain. The mouse genomic-human cDNA fusion effectively "humanizes" the beta-amyloid domain and facilitates the introduction of specific FAD mutations (shown as black asterisks) while leaving the remainder of mouse APP protein sequences unchanged. The human cDNA has been mutagenized to encode either the "Swedish"-FAD, "London"-FAD, "Swedish"/"London"-

- 65 -

FAD (shown in figure 2c), or "Swedish"-FAD APP713 mutations of APP (see also Fig. 10 and Table I).

Identification of targeted ES cell clones

After electroporation of each targeting vector (see Table I), ES cells were cultured for approximately 2 weeks in the presence of both positive (G418) and negative (FIAU) selection compounds. Four hundred G418/FIAU resistant ES cell colonies (clones) were then individually picked and cultured separately in 96 well culture dishes. The culture dishes were replica-plated, one set of copies was frozen to maintain the clones and the other replicate set was utilized for genetic analyses. From each well, DNA was extracted, digested with restriction enzyme, and analyzed by miniSouthern-blot analyses (see below). ES cell clones which appear to contain a targeted APP gene locus were thawed and expanded in culture. Gene-targeting was confirmed by Southern-blot analyses using DNA extracted from these expanded clones prior to introduction of the ES cell into the mouse germline (see below).

The mutagenesis of human cDNA's to encode the Swedish-FAD mutation also created a new Xba I (shown as X) restriction enzyme site (Figure 14). Incorporation of human FAD cDNA (shown in red) into the targeted m/hAPP gene locus thus changes the pattern of DNA fragments generated after digestion of this locus with Xba I. Using Southern-blot analyses, ES cell clones having the targeted m/hAPP gene can be distinguished from neomycin resistant ES cell clones having undesired random integrations of the targeting vector. The mouse exon 16 gene locus can be detected using a 3Kb Nco I (N) DNA fragment from intron 15 of the mouse APP gene as probe. Digestion of the mouse APP gene with Xba I generates an approximately 9 Kb DNA fragment whereas Xba I digestion of the targeted Swedish-FAD m/hAPP gene gives an approximately 5 Kb DNA

- 66 -

fragment when detected by Southern-blot analysis (see figure 14). This detection strategy applies to the Swedish-FAD m/hAPP, Swedish/London-FAD m/hAPP, and Swedish-FAD APP713 mutations.

Mini-Southern blot analysis identified 4 ES cell clones which appeared to contain the targeted Swedish-FAD m/h APP locus (data not shown).

These clones were expanded and subsequent Southern-blot analysis demonstrated that ES cell clones A79, A80, and B12 contain the Swedish FAD APP mutation while clone A72 did not (Figure 15). DNA extracted from ES cell pellets was examined by Southern-blot analysis using the restriction enzyme Xba I as described in Figure 14. A single ~9Kb DNA fragment is observed in non-targeted ES cells whereas targeted ES cell clones exhibit both the non-targeted allele (~9Kb fragment) and the FAD mutant allele giving rise to a ~5 Kb band. Transgenic mouse line ES5007 was derived from ES cell clone B12 (Table I). The remaining positive ES cell clones failed to establish germline transmission of the FAD mutation.

Initial miniSouthern-blot analyses identified five ES cell clones which appeared to contain Swedish/London FAD APP double mutation (data not shown). DNA extracted from pellets of expanded ES cell clones was examined by Southern-blot analysis using the restriction enzyme Xba I as described in Figure 14. This analysis confirmed that ES cell clones C82, C87, D25 and D92 contained the Swedish/London FAD m/hAPP double mutation while clones C52 and D49 did not. Transgenic mouse line ES5103 was derived from ES cell clone C87 (see Table I). The remaining positive ES cell clones failed to establish germline transmission of the FAD mutation. Confirmatory Southern-blot analyses identified multiple clones which carry the Swedish-FAD m/hAPP713 mutations (data not shown).

While identical in all other respects, the targeting vector encoding London-FAD m/hAPP does not carry the Xba I restriction site associated with the Swedish mutation. It was necessary, therefore, to identify restriction

- 67 -

enzymes which could distinguish between un-altered ES clones and those ES cell clones containing a targeted m/hAPP allele. The restriction enzymes Bcl I and BpM I were found to distinguish between DNA from a non-targeted ES cell clone (clone A1) and DNA from clone A21 which contains the Swedish-FAD m/hAPP gene locus (data not shown, see also figure 19). Bcl I and BpM I can be used to identify targeted clones derived from any of the aforementioned gene-targeting vectors.

Using restriction enzyme Bcl I, miniSouthern-blot analysis identified 6 ES cell clones which appeared to contain the London-FAD mutation (data not shown). Confirmatory Southern-blot analysis, using restriction enzyme BpM I, demonstrated that ES cell clones D12, D60, D70, D74, and D90 contained the London-FAD m/hAPP targeted locus while clone D45 did not (figure 18). After digestion with BpM I, three DNA fragments (~6 Kb, ~3.8 Kb, and ~2.2 Kb) are observed in non-targeted ES cells whereas targeted ES cell clones exhibited an additional ~4.8 Kb DNA fragment from the targeted allele (see A21 targeted for Swedish mutation. Transgenic mouse lines ES5401 and ES5403 were derived from ES cell clones D12 and D60 respectively (Table I). The remaining positive ES cell clones failed to establish germline transmission of the FAD mutation.

Germline-transmission of targeted m/hAPP genes

ES cells, confirmed to contain a targeted m/hAPP allele, were injected into the blastocoel cavity of a 3.5 day pre-implantation embryo (blastocyst). The injected blastocysts were then surgically reimplanted into pseudopregnant fosters and chimeras were born after approximately 17 days. The ES cells were derived from the 129SVEV inbred mouse strain which has a dominant agouti coat color gene. The blastocysts were derived from the C57BL/6 inbred mouse strain which carries a recessive black coat color gene. The coat color of chimeric mice whose cells are predominately derived from the ES

- 68 -

cells (designated as "high percentage chimeras") is mostly agouti with small patches of black. To establish germline transmission of the targeted APP gene, high percentage chimeric male mice were mated with either 129/SVEV inbred or black Swiss outbred females. The genotype of offspring from
5 ES5007, ES5103, ES5401 and ES5403 breeding pairs was determined by either Southern-blot or PCR analyses.

The Southern-blot analyses could distinguish between non-targeted, heterozygous, and homozygous progeny mice. The analyses utilized either Bcl I or Bpm I restriction enzyme, and the ~3.0Kb intron 15 DNA fragment
10 as probe (see Figure 14 for description of probe). A Southern-blot characterizing DNA from progeny of heterozygous ES5007 breeding pairs was performed. The technique can be applied to all the aforementioned transgenic lines. Bcl I digestion of non-transgenic (wt) mouse DNA and non-targeted ES cell DNA generated ~16 and ~8.5 Kb DNA fragments.
15 However, Bcl I digestion of heterozygous transgenic mouse DNA and targeted ES cell DNA generated ~16, ~8.5, and ~8.0 Kb DNA fragments. Digestion of homozygous mouse DNA with Bcl I liberated ~8.5 and ~8.0 Kb DNA fragments. Bpm I digestion of non-transgenic (wt) mouse DNA and non-targeted ES cell DNA generated ~6.0, ~3.8, and ~2.2 Kb DNA
20 fragments. Bpm I digestion of heterozygous transgenic mouse DNA and targeted ES cell DNA generated ~6.0, ~4.8, ~3.8, and ~2.2 Kb DNA fragments. Digestion of homozygous mouse DNA with Bpm I liberated ~6.0, ~4.8, and ~2.2Kb DNA fragments.

The genotype of offspring from ES5007, ES5103, ES5401 and ES5403
25 breeding pairs was also determined by PCR analyses using a combination of oligo pairs specific to human APP (H) and mouse APP (M) sequences. Like the Southern-blot technique, PCR analysis can distinguish between non-targeted, heterozygous, and homozygous progeny mice. As an internal standard, a 154 bp region of the mouse ribosomal subunit L32 gene is

- 69 -

amplified using the PCR oligo pair 6 and 7 (Figure 9). This control reaction was performed in each reaction. A 118 bp region specific to the mouse APP gene is amplified using the "M" oligo pair (oligos KC125 and KC132; Figure 9) and a 109 bp region specific to targeted m/hAPP gene is amplified using the "H" oligo pair (oligos KC125 and KC131; Figure 9). A PCR reaction using non-transgenic mouse DNA (wt) gives rise to a 118 bp fragment using the "M" oligo pair but no reaction product using the "H" oligo pair.

Conversely, a PCR reaction using DNA from transgenic mice homozygous (homoz.) for the targeted m/hAPP gene gives rise to a 109 bp fragment using the "H" oligo pair but no reaction product is observed using the "M" oligo pair. A PCR reaction using DNA from heterozygous transgenic mice (heter.) gives rise to both mouse and human PCR reaction products.

Messenger RNA (mRNA) expression in transgenic mouse brain

Analysis of APP mRNA composition in control mouse and ES5007 mouse brain has been determined using both Northern-blot and rtPCR analyses. RNA analyses have yet to be performed on the ES5103, ES5401, and ES5403 lines (Table II).

APP mRNA transcripts from control and ES5007 mouse brain were detected by Northern-blot analysis using an approximately 900 bp Nru I- Xho I fragment from pMTI-2385B (human APP cDNA) as probe. Mouse beta-actin mRNA was detected using a 430 bp mouse beta-actin cDNA probe (430 bp PCR product generated using oligos KC137 and KC138; see Figure 9) and served as an internal standard. mRNA from human brain (Clontech) served as a positive control.

mRNA from the Swedish-FAD m/hAPP gene was abundantly expressed in the brain from homozygous ES5007 mice. The amount of Swedish-FAD m/hAPP mRNA in ES5007 brain was determined by phosphoimage analysis and shown to be approximately 55% of the mAPP

- 70 -

mRNA levels observed in control mouse brain. In concordance, the APP mRNA levels in heterozygous ES5007 mouse brain were found to be approximately 75% of the level observed in control mouse brain.

The reverse transcriptase-PCR (rtPCR) technique was used to identify mouse APP and m/hAPP transcripts in mouse brain. Homozygous ES5007 mice were found to express mRNA exclusively from the targeted Swedish-FAD m/hAPP gene. No mRNA species containing sequences from mouse APP exons 16, 17, or 18 was detected in homozygotes. Heterozygous ES5007 mice were found to express mRNA transcripts from both mouse APP and Swedish-FAD m/hAPP alleles.

mRNA was purified from control and transgenic mouse brain and cDNAs were prepared using reverse transcriptase and oligonucleotide RA49 as primer. A 367 bp DNA fragment was amplified from mouse APP and m/hAPP cDNA by PCR using oligonucleotides KC56 and RA49 (Figure 9). Oligonucleotides KC56 and RA49 exhibit sequence identity with both mouse and human sequences. The mouse and human sequences were distinguished from each other by the presence of a Sty I restriction site in the human cDNA and the absence of the Sty I site in the mouse cDNA. Digestion of the 367 bp PCR product from m/hAPP cDNA generates two fragments (151 bp and 216 bp) while the PCR product from the mouse APP cDNA is not digested and remains unchanged at 367 bp.

APP mRNA from control mouse brain was amplified by rtPCR to generate a 367 bp DNA fragment that was resistant to Sty I digestion. rtPCR amplification of m/hAPP mRNA from the brain of homozygous ES5007 mice generated two fragments (151 bp and 216 bp) upon digestion by Sty I. No 367 bp DNA fragment remained, demonstrating that mouse APP cDNA was not present. All three DNA fragments (151 bp, 216 bp, and 367 bp) were observed after Sty I digestion of rtPCR product derived from heterozygous ES5007 brain transcripts. As a control, Sty I digestion of PCR

- 71 -

products from human APP cDNAs derived from human mRNA and mRNA from a HEK293 cell line expressing human APP generated the 151 and 216 bp DNA fragments. As expected, Sty I failed to digest the 367 bp PCR product derived from mouse brain mRNA.

5 m/hAPP protein expression in transgenic mouse brain

Swedish-FAD m/hAPP protein is expressed in the brain of ES5007 mice. MAb 286.8 specifically immunoprecipitates human APP and but not mouse APP. The epitope for MAb 286.8 has been determined to lie within the N-terminus of the human beta-amyloid domain (P. Graham et al. 1994, Pharma Report MRC 00116). The m/hAPP gene product could be specifically immunoprecipitated from a ES5007 brain homogenate using the monoclonal antibody (MAb) 286.8. APP moieties were then visualized by Western-blot analysis using MAb 22C11 as the detection antibody. MAb 22C11 can detect both mouse APP and m/hAPP. Therefore, if mouse APP was present after the immunoprecipitation it would have been detected by MAb 22C11. MAb 286.8 immunoprecipitated baculovirus derived human APP but did not recognize mouse APP in mouse brain homogenates.

The immunoprecipitations were performed using equal amounts of control mouse and ES5007 brain homogenates directly applied to the Western-blot. The relative intensities of the mouse APP and m/hAPP bands were equivalent.

Baculovirus derived human APP was directly applied to the Western-blot. An equal amount of human APP was detected after immunoprecipitation by MAb 286.8. It can be concluded, therefore, that MAb 286.8 efficiently immunoprecipitated human APP.

The expression of Swedish-FAD m/hAPP protein was further demonstrated by Western-blot analyses using additional detection antibodies. m/hAPP was immunoprecipitated from a homogenate of ES5007 brain using

- 72 -

human-specific MAb286.8. APP was then detected by Western-blot analysis using either the polyclonal antibody (PAb) 369 or MAb 6E10 for detection. MAb 6E10 is human APP specific and recognizes the human beta-amyloid domain. Again MAb 286.8 immunoprecipitates human APP, Swedish-FAD m/hAPP but does not immunoprecipitate mouse APP.

Swedish-FAD m/hAPP protein is expressed in the brain of homozygous ES5007 mice at approximately 87% of the level observed for mouse APP in non-transgenic mice. The relative expression values were determined in 3 independent Western-blot analyses using homogenates of brains from 4 homozygous ES5007 and 4 non-transgenic mice. The levels of m/hAPP protein expression ranged from 62% to 130% of control mouse APP depending on the protocol. In one experiment, APP was immunoprecipitated from equal amounts of brain homogenates from non-transgenic and homozygous ES5007 mice using PAb 369. For the other two Western-blot analyses, MAb 4G8 was used to immunoprecipitate APP. In all Western-blots, mouse APP and Swedish-FAD m/hAPP were visualized using MAb 22C11 as the detection antibody.

Processing of C-terminal domain of APP

The Swedish FAD mutation significantly altered the proteolytic processing of the APP resulting in a change in the C-terminal fragments of APP. The observed changes in processing was consistent with a predominant usage of the beta-secretase site over the alpha-secretase site.

Membrane preparations from brain homogenates were solubilized by detergents and APP holoprotein and C-terminal fragments were immunoprecipitated using MAb 4G8. Mouse APP and m/hAPP holoproteins were detected by Western-blot analysis using MAb 22C11. The C-terminal fragments from both mouse APP and m/hAPP were detected using PAb369 while C-terminal fragments derived exclusively from m/hAPP were detected

- 73 -

using human specific MAb 6E10. The expression level of m/hAPP in homozygous ES5007 (Swed-homoz) brain was found to be approximately 62% of the level observed for mouse APP.

Under normal conditions, the proteolytic processing of mouse APP resulted in the generation of 5 C-terminal fragments. This contrasts with the pattern observed with Swedish-FAD m/hAPP where only the two largest C-terminal fragments were observed. The second largest C-terminal fragment (fragment 2) co-migrated with the LEC100 standard. The electrophoretic mobility of LEC100 was expected to closely resemble that of the C-terminal fragment released after the cleavage by beta-secretase. LEC100 consists of amino acid residues Leu, Gly, and Met juxtaposed with the beta-amyloid, transmembrane, and cytoplasmic domains of APP. spLEC100 (sp designates APP signal peptide, see below) was stably expressed in HEK293 cells (obtained from Sandra Reuter), a membrane homogenate prepared, and an aliquot was applied to the gel. In cells, LEC100 is generated after the signal peptide (sp) is proteolytically removed from spLEC100 during protein translation.

For other aspects of the nucleic acids, polypeptides, antibodies, etc., reference is made to standard textbooks of molecular biology, protein science, and immunology. See, e.g., Davis et al. (1986), *Basic Methods in Molecular Biology*, Elsevir Sciences Publishing, Inc., New York; Hames et al. (1985), *Nucleic Acid Hybridization*, IL Press, *Molecular Cloning*, Sambrook et al.; *Current Protocols in Molecular Biology*, Edited by F.M. Ausubel et al., John Wiley & Sons, Inc; *Current Protocols in Human Genetics*, Edited by Nicholas C. Dracopoli et al., John Wiley & Sons, Inc.; *Current Protocols in Protein Science*; Edited by John E. Coligan et al., John Wiley & Sons, Inc.; *Current Protocols in Immunology*; Edited by John E. Coligan et al., John Wiley & Sons, Inc.

- 74 -

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The preceding preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The entire disclosure of all applications, patents and publications, cited below are hereby incorporated by reference.

From the foregoing description, one skilled in the art can easily ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

100

- 75 -

Tables

Table I

Transgenic line	ES cell clone	Targeting vector	Gene Product	mAPP mutations
ES5007	B12	pMTI-2398	Swedish-FAD m/h APP	KM(670,671)NL; G(676)R; F(681)T, R(684)H
ES5103	C87	pMTI-2454	Swedish/Londo n-FAD m/h APP	KM(670,671)NL; V(717)I; G(676)R, F(681)T, R(684)H
ES5215	A54	pMTI-2455	Swedish-FAD m/hAPP713	KM(670,671)NL; T(714)stop; G(676)R, F(681)T, R(684)H
ES5401	D12	pMTI-2453	London-FAD m/h APP	V(717)I; G(676)R, F(681)T, R(684)H
ES5403	D60	pMTI-2453	London-FAD m/h APP	V(717)I; G(676)R, F(681)T, R(684)H

- 76 -

Table II

Transgenic Line	Germline- Transmission	m/h APP mRNA	m/h APP Protein	Altered C- Terminal Processing
ES5007	yes	yes	yes	yes
ES5103	yes	n.d.	yes	yes
ES5401	yes	n.d.	yes	n.d.
ES5403	yes	n.d.	yes	n.d.

n.d.: not determined

- 77 -

- G.R. Askew, T. Doetschman and J.B. Lingrel, *Mol. Cell Biol.*, 13:4115-24 (1993)
- C. Bonnerot, G. Grimmer, P. Briand and J.F. Nicolas, *Proc. Natl. Acad. Sci. USA*, 87:6331-6335 (1990)
- A. Bradley, in E.J. Robertson (ed.), *Production and analysis of chimeric mice*, IRL Press, Oxford and Washington, DC, pp. 113-151 (1987)
- R.L. Brinster, J.M. Allen, R.R. Behringer, R.E. Gelinas and R.D. Palmiter, *Proc. Natl. Acad. Sci. USA*, 85:836-840 (1988)
- X.D. Cai, T.E. Golde and S.G. Younkin, *Science*, 259:514-516 (1993)
- M. Citron, C. Vigo-Pelfrey, D.B. Teplow, C. Miller, D. Schenk, J. Johnston, B. Wilblad, N. Venizelos, L. Lannfelt and D.J. Selkoe, *Proc. Natl. Acad. Sci. USA*, 91:11993-11997 (1994)
- T. Dyrks, E. Dyrks, C.L. Masters and K. Beyreuther, *FEBS Lett.*, 324:231-236 (1993)
- D. Games, D. Adams, R. Alessandrini, R. Barbour, P. Berthelette, C. Blackwell, T. Carr, J. Clemens, T. Donaldson, F. Gillespie, et al., *Nature*, 373:523-527 (1995)
- S.A.e.a. Gravina (1995)
- F. Grosveld, G.B. van Assendelft, D.R. Greaves and G. Kollias, *Cell*, 51:975-985 (1987)
- P. Hasty, R. Ramirez-Solis, R. Krumlauf and A. Bradley, *Nature*, 350:243-246 (1991)
- L.S. Higgins et al. (1995)
- L.S. Higgins, D.M. Holtzman, J. Rabni, W.C. Mobley and B. Cordell, *Ann. Neurol.*, 35:598-607 (1994)
- J.Kang and B. Muller-Hill, *Biochem. Biophys. Res. Commun.*, 166:1192-1200 (1990)
- F.M. La Ferla, B.T. Tinkle, C.J. Bieberich, C.C. Haudenschild and G. Jay, *Nat. Genet.*, 9:21-30 (1995)

- 78 -

- F.M. La Ferla, D.A. Kappel Hall, L. Ngo and G. Jay, In manuscript (1995)
- B.T. Lamb, *Nat. Genet.*, 9:4-6 (1995)
- B.T. Lamb, S.S. Sisodia, A.M. Lawler, H.H. Slunt, C.A. Kitt, W.G. Kearns, P.L. Pearson, D.L. Price and J.D. Gearhart, *Nat. Genet.*, 5:22-30 (1993)
- H.G. Lemaire, J.M. Salbaum, G. Multhaup, J. Kang, R.M. Bayney, A. Unterbeck, K. Beyreuther and B. Muller-Hill, *Nucleic Acids Res.*, 17:517-522 (1989)
- B.E. Pearson and T.K. Choi, *Proc. Natl. Acad. Sci. USA*, 90:10578-10582 (1993)
- E.J. Robertson, in E.J. Robertson (ed.), *Embryo-derived stem cell lines*, IRL Press, Oxford and Washinton, DC, pp. 71-112 (1987)
- M. Rubinstein, M.A. Japon and M.J. Low, *Nucleic Acids Res.*, 21:2613-2617 (1993)
- W. S. Simonet, N. Bucay, S.J. Lauer and J.M. Taylor, *J. Biol. Chem.*, 268:8221-8229 (1993)
- A. Stacey, A. Schnieke, J. McWhir, J. Cooper, A. Colman and D.W. Melton, *Mol. Cell Biol.*, 14:1009-1016 (1994)
- N. Suzuki, T.T. Cheung, X.D. Cai, A. Odaka, L. Otvos, Jr., C. Eckman, T.E. Golde and S.G. Younkin, *Science*, 264:1336-1340 (1994)

- 79 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: WIRAK, Dana O.
- (ii) TITLE OF INVENTION: METHOD OF INTRODUCING MODIFICATIONS INTO A GENE
- (iii) NUMBER OF SEQUENCES: 36
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Bayer Corporation
 - (B) STREET: 400 Morgan Lane
 - (C) CITY: West Haven
 - (D) STATE: CT
 - (E) COUNTRY: US
 - (F) ZIP: 06516-4175
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/698,360
 - (B) FILING DATE: 15-AUG-1996
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Jones, Huw R.
 - (B) REGISTRATION NUMBER: 33,916
 - (C) REFERENCE/DOCKET NUMBER: WH 5009-PCT
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (203) 812-2317
 - (B) TELEFAX: (203) 812-5492

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 38 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Adaptor"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

- 80 -

TCGACGACTT AAGTTGATAT CCACCATGGT GACGCGTT

38

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Adaptor"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TCGAGTGAGA TCTTAAGGCC TGG

23

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 58 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Adaptor"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TCGACAAGGC GCGCGTTTAA ACAAGCGGC CGCTTGGCGC GCCTTTTGTT TAAACTTG

58

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Primer"

(iii) HYPOTHETICAL: NO

- 81 -

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CCTCGGCCTT TGGTGTGTGT TTTATGACAT GACCCCCTTG A

41

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CACCCTGTTG TCAATGCCTC TGGGTTTCCG CCAGTTTCG

39

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGATGGGTAG TGAAGCA

17

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- 82 -

(A) DESCRIPTION: /desc = "Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GTGAAGATGG ATGCAGAATT C

21

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTTCTGGGCT GACAAACATC

20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GATGGCGGAC TTCAAATCCT G

21

- 83 -

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTAGACACTC

10

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 16 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ACTTTGTGTT TGACGC

16

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

- 84 -

GATGATGAAC TTCATATCCT G

21

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CAGTTTTTTGA TGGCGG

16

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTTTGAGACC TTCAACACCC

20

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Primer"

(iii) HYPOTHETICAL: NO

- 85 -

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAAGGAAGGC TGGAAAAGAG CC

22

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 770 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met	Leu	Pro	Gly	Leu	Ala	Leu	Leu	Leu	Leu	Ala	Ala	Trp	Thr	Ala	Arg	1	5	10	15
Ala	Leu	Glu	Val	Pro	Thr	Asp	Gly	Asn	Ala	Gly	Leu	Leu	Ala	Glu	Pro	20	25	30	
Gln	Ile	Ala	Met	Phe	Cys	Gly	Arg	Leu	Asn	Met	His	Met	Asn	Val	Gln	35	40	45	
Asn	Gly	Lys	Trp	Asp	Ser	Asp	Pro	Ser	Gly	Thr	Lys	Thr	Cys	Ile	Asp	50	55	60	
Thr	Lys	Glu	Gly	Ile	Leu	Gln	Tyr	Cys	Gln	Glu	Val	Tyr	Pro	Glu	Leu	65	70	75	80
Gln	Ile	Thr	Asn	Val	Val	Glu	Ala	Asn	Gln	Pro	Val	Thr	Ile	Gln	Asn	85	90	95	
Trp	Cys	Lys	Arg	Gly	Arg	Lys	Gln	Cys	Lys	Thr	His	Pro	His	Phe	Val	100	105	110	
Ile	Pro	Tyr	Arg	Cys	Leu	Val	Gly	Glu	Phe	Val	Ser	Asp	Ala	Leu	Leu	115	120	125	
Val	Pro	Asp	Lys	Cys	Lys	Phe	Leu	His	Gln	Glu	Arg	Met	Asp	Val	Cys	130	135	140	
Glu	Thr	His	Leu	His	Trp	His	Thr	Val	Ala	Lys	Glu	Thr	Cys	Ser	Glu	145	150	155	160
Lys	Ser	Thr	Asn	Leu	His	Asp	Tyr	Gly	Met	Leu	Leu	Pro	Cys	Gly	Ile	165	170	175	

- 86 -

Asp Lys Phe Arg Gly Val Glu Phe Val Cys Cys Pro Leu Ala Glu Glu
 180 185 190
 Ser Asp Asn Val Asp Ser Ala Asp Ala Glu Glu Asp Asp Ser Asp Val
 195 200 205
 Trp Trp Gly Gly Ala Asp Thr Asp Tyr Ala Asp Gly Ser Glu Asp Lys
 210 215 220
 Val Val Glu Val Ala Glu Glu Glu Glu Val Ala Glu Val Glu Glu Glu
 225 230 235 240
 Glu Ala Asp Asp Asp Glu Asp Asp Glu Asp Gly Asp Glu Val Glu Glu
 245 250 255
 Glu Ala Glu Glu Pro Tyr Glu Glu Ala Thr Glu Arg Thr Thr Ser Ile
 260 265 270
 Ala Thr Thr Thr Thr Thr Thr Thr Glu Ser Val Glu Glu Val Val Arg
 275 280 285
 Glu Val Cys Ser Glu Gln Ala Glu Thr Gly Pro Cys Arg Ala Met Ile
 290 295 300
 Ser Arg Trp Tyr Phe Asp Val Thr Glu Gly Lys Cys Ala Pro Phe Phe
 305 310 315 320
 Tyr Gly Gly Cys Gly Gly Asn Arg Asn Asn Phe Asp Thr Glu Glu Tyr
 325 330 335
 Cys Met Ala Val Cys Gly Ser Ala Met Ser Gln Ser Leu Leu Lys Thr
 340 345 350
 Thr Gln Glu Pro Leu Ala Arg Asp Pro Val Lys Leu Pro Thr Thr Ala
 355 360 365
 Ala Ser Thr Pro Asp Ala Val Asp Lys Tyr Leu Glu Thr Pro Gly Asp
 370 375 380
 Glu Asn Glu His Ala His Phe Gln Lys Ala Lys Glu Arg Leu Glu Ala
 385 390 395 400
 Lys His Arg Glu Arg Met Ser Gln Val Met Arg Glu Trp Glu Glu Ala
 405 410 415
 Glu Arg Gln Ala Lys Asn Leu Pro Lys Ala Asp Lys Lys Ala Val Ile
 420 425 430
 Gln His Phe Gln Glu Lys Val Glu Ser Leu Glu Gln Glu Ala Ala Asn
 435 440 445

- 87 -

Glu Arg Gln Gln Leu Val Glu Thr His Met Ala Arg Val Glu Ala Met
 450 455 460
 Leu Asn Asp Arg Arg Arg Leu Ala Leu Glu Asn Tyr Ile Thr Ala Leu
 465 470 475 480
 Gln Ala Val Pro Pro Arg Pro Arg His Val Phe Asn Met Leu Lys Lys
 485 490 495
 Tyr Val Arg Ala Glu Gln Lys Asp Arg Gln His Thr Leu Lys His Phe
 500 505 510
 Glu His Val Arg Met Val Asp Pro Lys Lys Ala Ala Gln Ile Arg Ser
 515 520 525
 Gln Val Met Thr His Leu Arg Val Ile Tyr Glu Arg Met Asn Gln Ser
 530 535 540
 Leu Ser Leu Leu Tyr Asn Val Pro Ala Val Ala Glu Glu Ile Gln Asp
 545 550 555 560
 Glu Val Asp Glu Leu Leu Gln Lys Glu Gln Asn Tyr Ser Asp Asp Val
 565 570 575
 Leu Ala Asn Met Ile Ser Glu Pro Arg Ile Ser Tyr Gly Asn Asp Ala
 580 585 590
 Leu Met Pro Ser Leu Thr Glu Thr Lys Thr Thr Val Glu Leu Leu Pro
 595 600 605
 Val Asn Gly Glu Phe Ser Leu Asp Asp Leu Gln Pro Trp His Ser Phe
 610 615 620
 Gly Ala Asp Ser Val Pro Ala Asn Thr Glu Asn Glu Val Glu Pro Val
 625 630 635 640
 Asp Ala Arg Pro Ala Ala Asp Arg Gly Leu Thr Thr Arg Pro Gly Ser
 645 650 655
 Gly Leu Thr Asn Ile Lys Thr Glu Glu Ile Ser Glu Val Lys Met Asp
 660 665 670
 Ala Glu Phe Arg His Asp Ser Gly Tyr Glu Val His His Gln Lys Leu
 675 680 685
 Val Phe Phe Ala Glu Asp Val Gly Ser Asn Lys Gly Ala Ile Ile Gly
 690 695 700
 Leu Met Val Gly Gly Val Val Ile Ala Thr Val Ile Val Ile Thr Leu
 705 710 715 720
 Val Met Leu Lys Lys Lys Gln Tyr Thr Ser Ile His His Gly Val Val
 725 730 735

- 88 -

Glu Val Asp Ala Ala Val Thr Pro Glu Glu Arg His Leu Ser Lys Met
740 745 750

Gln Gln Asn Gly Tyr Glu Asn Pro Thr Tyr Lys Phe Phe Glu Gln Met
755 760 765

Gln Asn
770

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11992 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Mus musculus

(ix) FEATURE:

- (A) NAME/KEY: mat_peptide
- (B) LOCATION: 6541..6639

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

NNNAAGCTTN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	60
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	120
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	180
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	240
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	300
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	360
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	420
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	480
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	540
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	600
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	660
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	720

[illegible]

- 91 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4140
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4200
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4260
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4320
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4380
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4440
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4500
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4560
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4620
NNNNNNNNNN	GGTACCNNNC	CATGGNNNNN	NNNNNNNNNN	NNNAAGCTTN	NNNNNNNNNN	4680
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4740
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4800
NNNNNNNNNN	NCATATGNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4860
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4920
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4980
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	5040
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	5100
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	5160
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	5220
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	5280
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNGAA	TTCNNNNNNN	NNNNNNNNNN	5340
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	5400
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	5460
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	5520
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNG	AATTCNNNNN	NNNNNNNNNN	5580
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	5640
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	5700
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	5760

- 93 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	7500
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	7560
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	7620
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	7680
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	7740
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	7800
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	7860
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	7920
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	7980
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8040
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8100
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8160
GATCCNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8220
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8280
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8340
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8400
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8460
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8520
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8580
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8640
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8700
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8760
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8820
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8880
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8940
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9000
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9060
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9120

- 94 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9180
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9240
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9300
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9360
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9420
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9480
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9540
NNNNNNNNNN	NNNNNNNNNN	GGCGCCNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9600
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9660
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9720
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9780
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9840
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9900
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9960
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10020
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10080
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10140
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	CCATGGNNNN	NNNNNNNNNN	NNNNNNNNNN	10200
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10260
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10320
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10380
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10440
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNTCTAGANN	10500
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNAGATCT	NNNNNNNNNN	10560
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10620
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10680
NAAGCTTNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10740
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNAG	ATCTNNNNNN	NNNNNNNNNN	10800

- 95 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10860
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10920
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10980
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11040
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11100
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNGA	TATCNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11160
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11220
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11280
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11340
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11400
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11460
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11520
NNNNNNNNNN	NNNNNNNNNN	NNCATATGNN	NGAATTTCNN	NNNNNNNNNN	NNNNNNNNNN	11580
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11640
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11700
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNGT	ATACNNNNNN	11760
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NGGTACCNNN	NNNNNNNNNN	11820
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11880
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNCAGC	11940
TGNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNCCAT	GG	11992

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12814 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Targetting vector"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

- 96 -

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(C) INDIVIDUAL ISOLATE: Swedish-FAD APP

(ix) FEATURE:

(A) NAME/KEY: mat_peptide

(B) LOCATION: 1932..2276

(D) OTHER INFORMATION: /standard_name= "Swedish-FAD APP"

(ix) FEATURE:

(A) NAME/KEY: mat_peptide

(B) LOCATION: 5360..6160

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GAGCTCCACC GCGGTGGCGG CCGCTCTGAC CATGGNNNNN NNNNNNNNNN NNNAAGCTTN	60
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	120
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	180
NNNNNNNNNN NNNNNNNNNN NCATATGNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	240
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	300
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	360
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	420
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	480
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	540
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	600
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	660
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNGAA TTCNNNNNNN	720
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	780
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	840
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	900
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNNG AATTCNNNNN	960
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	1020
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	1080
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	1140
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	1200
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNGAATTCN	1260

- 97 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1320
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1380
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1440
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1500
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1560
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1620
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1680
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1740
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1800
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1860
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	1920
NNNNNNNNNG	TTCTGGGCTG	ACAAACATCA	AGACGGAAGA	GATCTCTGAA	GTGAATCTAG	1980
ATGCAGAATT	CCGACATGAC	TCAGGATATG	AAGTTCATCA	TCAAAAATTG	GTGTTCTTTG	2040
CAGAAGATGT	GGGTTCAAAC	AAAGGTGCAA	TCATTGGACT	CATGGTGGGC	GGTGTGTGCA	2100
TAGCGACAGT	GATCGTCATC	ACCTTGGTGA	TGCTGAAGAA	GAAACAGTAC	ACATCCATTC	2160
ATCATGGTGT	GGTGGAGGTT	GACGCCGCTG	TCACCCAGA	GGAGCGCCAC	CTGTCCAAGA	2220
TGCAGCAGAA	CGGCTACGAA	AATCCAACCT	ACAAGTTCTT	TGAGCAGATG	CAGAACTAGA	2280
CCCCCGCCAC	AGCAGCCTCT	GAAGTTGGAC	AGCAAAACCA	TTGCTTCACT	ACCCATCGGT	2340
GTCCATTTAT	AGAATAATGT	GGGAAGAAAC	AAACCCGTTT	TATGATTTAC	TCATTATCGC	2400
CTTTTGACAG	CTGTGCTGTA	ACACAAGTAG	ATGCCTGAAC	TTGAATTAAT	CCACACATCA	2460
GTAATGTATT	CTATCTCTCT	TTACATTTTG	GTCTCTATAC	TACATTATTA	ATGGGTTTTG	2520
TGTACTGTAA	AGAATTTAGC	TGTATCAAAC	TAGTGCATGA	ATAGATTCTC	TCCTGATTAT	2580
TTATCACATA	GCCCCTTAGC	CAGTTGTATA	TTATTCTTGT	GGTTTGTGAC	CCAATTAAGT	2640
CCTACTTTAC	ATATGCTTTA	AGAATCGATG	GGGGATGCTT	CATGTGAACG	TGGGAGTTCA	2700
GCTGCTTCTC	TTGCCTAAGT	ATTCCTTTCC	TGATCACTAT	GCATTTTAAA	GTAAACATT	2760
TTTAAGTATT	TCAGATGCTT	TAGAGAGATT	TTTTTTCCAT	GACTGCATTT	TACTGTACAG	2820
ATTGCTGCTT	CTGCTATATT	TGTGATATAG	GAATTAAGAG	GATACACACG	TTTGTTTCTT	2880
CGTGCCTGTT	TTATGTGCAC	ACATTAGGCA	TTGAGACTTC	AAGCTTTTCT	TTTTTTGTCC	2940

- 98 -

ACGTATCTTT	GGGTCTTTGA	TAAAGAAAAG	AATCCCTGTT	CATTGTAAGC	ACTTTTACGG	3000
GGCGGGTGGG	GAGGGGTGCT	CTGCTGGTCT	TCAATTACCA	AGAATTCTCC	AAAACAATTT	3060
TCTGCAGGAT	GATTGTACAG	AATCATTGCT	TATGACATGA	TCGCTTTCTA	CACTGTATTA	3120
CATAAATAAA	TTAAATAAAA	TAACCCCGGG	CAAGACTTTT	CTTTGAAGGA	TGACTACAGA	3180
CATTAAATAA	TCGAAGTAAT	TTTGGGTGGG	GAGAAGAGGC	AGATTCAATT	TTCTTTAACC	3240
AGTCTGAAGT	TTCATTTATG	ATACAAAAGA	AGATGAAAAT	GGAAGTGGCA	ATATAAGGGG	3300
ATGAGGAAGG	CATGCCTGGA	CAAACCCTTC	TTTTAAGATG	TGTCTTCAAT	TTGTATAAAA	3360
TGGTGTTTTT	ATGTAAATAA	ATACATTCTT	GGAGGAGCCA	CATTGTGCTG	GTGTGAATGA	3420
TTCCATAGTA	ACAATCTTGA	CCATTTACTG	ACGTACAGAC	CAGTGAGAAG	TCTTCGCATG	3480
TTGGGTACCC	ACACCTGTTG	TGTCTTAATT	GCAAGTCTGA	GTAGGAAGTT	GGGGCCAACA	3540
TGTGTCTCCC	AGTGCTGGGA	AAATATTTCA	TAGACCTAAT	TTACAGTCTT	TACTTGATCT	3600
AAAACATTTT	GCTGCCATAT	TTTGGCCCTC	AAGTTTGTCC	CAAATGAGAG	ACAAAGGGAA	3660
AAGTTCCAGG	GAAATAAAAA	TTAAGACAGC	TGATTATCTG	TAAAGCATGG	TTTCTCATCC	3720
TGAACGCTAC	TAACATTTTG	CAGGGAATAA	TTCCTTGTTG	AAGGGAGTTG	TCCTGACCAG	3780
TGTAGGATAT	TTATTTATTT	TATTTATGTT	TTTTGAGACG	GAGTCTCGCT	CTGTCACCCA	3840
GGCTGGAGTG	CAGTGGCACA	ATCTCGGCTC	ACTGCAAGCT	CCGCCTCCCG	GGTTCACGCC	3900
ATTCTCCTGC	CTCAGCCTCC	TGAATAGCTG	GGACTCTAGG	TGCCCCGCCAC	CACGCCCGGC	3960
TAATTTTTTG	TATTTTTAGT	AGAGACGGGG	TTTCACCGTG	TTAGCCAGGA	CAGTCTTGTT	4020
CTCCTGACCT	CGTGATCTGC	CTGCCTCGGC	CTCCCAAAGT	GCTGAGATTA	CAGGCGTGCA	4080
AGCCGCGCCC	AGCCAGTGCT	CTCCTTTTAA	AAGTAGCCCA	TTGGCTGGGC	GCAGTGGCTC	4140
ACGCCTGTAA	TCCCAGCACT	TTGGGAGGCT	GAGGCGGGTG	GATCACGAGG	TCAGGAGATC	4200
AAGAATATCC	TGGCCAATAT	GGTGAAACCC	CATCTCTACT	AAAAATACAA	AAAAAAAAAA	4260
AAAAAAAAAA	AGGCCGGGCA	TGGTGGCGGG	CGCTTGTAGT	CCCAGCTACT	CAGGAGGCTG	4320
AGGCAGGAGA	ATGGTGTGCA	CCTGGGAGGC	GGAGGTTGCA	GTGAGCTGAG	ATCGCGCCAC	4380
TGCACTCCAG	CCTGGGAGAC	AGAGCGAGAC	TCCGTCTCAA	TAAATAAATA	AATAAATAAA	4440
TAAAAGGAGG	GCCTGGCACG	AATGACATGC	AGGGAAGGCA	GTGAGCAGGT	GGAGGTCCCT	4500
GTA CTGTTG	TGGTGCCTTA	TCTACCAGGC	GGTTGAGTTG	ACGTCTTTGT	GGACAGAATT	4560
CGAGCTCGGT	ACCCGGGGAT	CCTCTAGAGT	CGACCTTAAG	GATATCCTTA	AGGTGCACGG	4620

- 99 -

TATCGATAAG	CTTGGGCTTG	AACATCGAGC	GCCAGGGCTC	CGTAAAGCTA	CTAGAGCACA	4680
GGCGGTGCCC	CAACGTCCTG	GGGCCTCTCC	ACTAATAACG	GCTACTTCCA	ATTGATTGGA	4740
CGCGCCATCT	TGCCTGCCTT	ATGCATATTC	AGCGGTGAAC	TGAATATTCA	TGAACGAGGC	4800
CCGTCCCGTC	CCTCCCTCCT	TCCCCCACC	CCCGGAACCC	GCTCCGGAGG	ACCCGAAGGG	4860
CCCCGCCTTC	ATTACCGATG	CGTAGGACAA	ACCATTTTCC	CGATGTGTGT	GGGGGGATAC	4920
TAATGAGAGA	CTTTAGCTGA	AAAATGAGCC	TGAACTCCGA	AGCTGAGTAA	AAATGGCCTA	4980
ACTTTATCCT	CCGTTCTGTA	AGTCCTCGGT	TTGAGTGCAC	GGGAAACCCG	AAAGGAGGAC	5040
GACAGGACCA	GGACATTCTC	CTCCTCCTGT	CGCGTCAGAA	AGAACACCCA	ACCAGGGAGC	5100
CGGAGCCCTA	GCGTCAACAA	CTCCGCCGCG	CGCGCTCCGT	GTAGGCCGGT	GCGGGCGGCC	5160
CCGTAGCGCA	AGGGAGGGCG	GGAAAGGAAG	GGGCGGGACA	CAAGGGCGAA	TCTATAAAGG	5220
GCGTCACTCA	GCCAGTTCTC	TCCTCAGAAG	CGCCGAGAGC	GCGACCGGGA	CGGTTGGAGA	5280
AGAAGGTGGC	TCCCGGAAGG	GGGAGAGACA	AACTGCCGTA	ACCTCTGCCG	TTCAGGATCA	5340
TCGAATTCCT	GCAGCCAATA	TGGGATCGGC	CATTGAACAA	GATGGATTGC	ACGCAGGTTC	5400
TCCGGCCGCT	TGGGTGGAGA	GGCTATTGCG	CTATGACTGG	GCACAACAGA	CAATCGGCTG	5460
CTCTGATGCC	GCCGTGTTCC	GGCTGTCAGC	GCAGGGGCGC	CCGTTTCTTT	TTGTCAAGAC	5520
CGACCTGTCC	GGTGCCCTGA	ATGAACTGCA	GGACGAGGCA	GCGCGGCTAT	CGTGGCTGGC	5580
CACGACGGGC	GTTCTTGCG	CAGCTGTGCT	CGACGTTGTC	ACTGAAGCGG	GAAGGGACTG	5640
GCTGCTATTG	GGCGAAGTGC	CGGGGCAGGA	TCTCCTGTCA	TCTCACCTTG	CTCCTGCCGA	5700
GAAAGTATCC	ATCATGGCTG	ATGCAATGCG	GCGGCTGCAT	ACGCTTGATC	CGGCTACCTG	5760
CCCATTGAC	CACCAAGCGA	AACATCGCAT	CGAGCGAGCA	CGTACTCGGA	TGGAAGCCGG	5820
TCTTGTCGAT	CAGGATGATC	TGGACGAAGA	GCATCAGGGG	CTCGCGCCAG	CCGAACTGTT	5880
CGCCAGGCTC	AAGGCGCGCA	TGCCCGACGG	CGAGGATCTC	GTCGTGACCC	ATGGCGATGC	5940
CTGCTTGCCG	AATATCATGG	TGGAAAATGG	CCGCTTTTCT	GGATTCATCG	ACTGTGGCCG	6000
GCTGGGTGTG	GCGGACCGCT	ATCAGGACAT	AGCGTTGGCT	ACCCGTGATA	TTGCTGAAGA	6060
GCTTGGCGGC	GAATGGGCTG	ACCGCTTCCT	CGTGCTTTAC	GGTATCGCCG	CTCCCGATTG	6120
GCAGCGCATC	GCCTTCTATC	GCCTTCTTGA	CGAGTTCTTC	TGAGGGGATC	AATTCTCTAG	6180
AGCTCGCTGA	TCAGCCTCGA	CTGTGCCTTC	TAGTTGCCAG	CCATCTGTTG	TTTGCCCCTC	6240
CCCCGTGCCT	TCCTTGACCC	TGGAAGGTGC	CACTCCCACT	GTCCTTTCCT	AATAAAATGA	6300

- 101 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8040
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8100
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8160
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8220
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8280
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8340
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8400
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8460
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8520
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8580
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8640
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8700
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8760
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8820
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8880
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	8940
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9000
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9060
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9120
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9180
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9240
NNNNNNNNNN	NNNNNNNNNN	GGCGCCNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9300
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9360
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9420
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9480
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9540
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9600
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9660

- 102 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9720
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9780
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9840
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NCCATGGTCT	AGAACTAGTG	GATCCCCCGG	9900
GCTGCAGGAA	TTCGATATCA	AGCTTATCGA	TACCGTCGAC	CTCGAGGGGG	GGCCCCGTAC	9960
CCAATTCGCC	CTATAGTGAG	TCGTATTACG	CGCGCTCACT	GGCCGTCGTT	TTACAACGTC	10020
GTGACTGGGA	AAACCCTGGC	GTTACCCAAC	TTAATCGCCT	TGCAGCACAT	CCCCCTTTTCG	10080
CCAGCTGGCG	TAATAGCGAA	GAGGCCCGCA	CCGATCGCCC	TTCCAACAG	TTGCGCAGCC	10140
TGAATGGCGA	ATGGGACGCG	CCCTGTAGCG	GCGCATTAA	CGCGGCGGGT	GTGGTGGTTA	10200
CGCGCAGCGT	GACCGCTACA	CTTGCCAGCG	CCCTAGCGCC	CGCTCCTTTC	GCTTTCCTTC	10260
CTTCCTTTCT	CGCCACGTTT	GCCGGCTTTC	CCCGTCAAGC	TCTAAATCGG	GGGCTCCCTT	10320
TAGGGTTCCG	ATTTAGTGCT	TTACGGCACC	TCGACCCCAA	AAAACCTGAT	TAGGGTGATG	10380
GTTACGCTAG	TGGGCCATCG	CCCTGATAGA	CGGTTTTTCG	CCCTTTGACG	TTGGAGTCCA	10440
CGTTCCTTAA	TAGTGGACTC	TTGTTCCAAA	CTGGAACAAC	ACTCAACCCT	ATCTCGGTCT	10500
ATTCTTTTGA	TTTATAAGGG	ATTTTGCCGA	TTTCGGCCTA	TTGGTTAAAA	AATGAGCTGA	10560
TTTAACAAAA	ATTTAACGCG	AATTTTAACA	AAATATTAAC	GCTTACAATT	TAGGTGGCAC	10620
TTTTTCGGGA	AATGTGCGCG	GAACCCCTAT	TTGTTTATTT	TTCTAAATAC	ATTCAAATAT	10680
GTATCCGCTC	ATGAGACAAT	AACCCTGATA	AATGCTTCAA	TAATATTGAA	AAAGGAAGAG	10740
TATGAGTATT	CAACATTTCC	GTGTCGCCCT	TATTCCTTTT	TTTGCGGCAT	TTTGCCTTCC	10800
TGTTTTTGCT	CACCCAGAAA	CGCTGGTGAA	AGTAAAAGAT	GCTGAAGATC	AGTTGGGTGC	10860
ACGAGTGGGT	TACATCGAAC	TGGATCTCAA	CAGCGGTAAG	ATCCTTGAGA	GTTTTCGCCC	10920
CGAAGAACGT	TTTCCAATGA	TGAGCACTTT	TAAAGTTCTG	CTATGTGGCG	CGGTATTATC	10980
CCGTATTGAC	GCCGGGCAAG	AGCAACTCGG	TCGCCGCATA	CACTATTCTC	AGAATGACTT	11040
GGTTGAGTAC	TCACCAGTCA	CAGAAAAGCA	TCTTACGGAT	GGCATGACAG	TAAGAGAATT	11100
ATGCAGTGCT	GCCATAACCA	TGAGTGATAA	CACTGCGGCC	AACTTACTTC	TGACAACGAT	11160
CGGAGGACCG	AAGGAGCTAA	CCGCTTTTTT	GCACAACATG	GGGGATCATG	TAACTCGCCT	11220
TGATCGTTGG	GAACCGGAGC	TGAATGAAGC	CATACCAAAC	GACGAGCGTG	ACACCACGAT	11280
GCCTGTAGCA	ATGGCAACAA	CGTTGCGCAA	ACTATTAAC	GGCGAACTAC	TTACTCTAGC	11340

- 103 -

TTCCCGGCAA CAATTAATAG ACTGGATGGA GCGGATAAAA GTTGCAGGAC CACTTCTGCG	11400
CTCGGCCCTT CCGGCTGGCT GGTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGGTC	11460
TCGCGGTATC ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA	11520
CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC	11580
CTCACTGATT AAGCATTGGT AACTGTCAGA CCAAGTTTAC TCATATATAC TTTAGATTGA	11640
TTTAAAACTT CATTTTTAAT TAAAAGGAT CTAGGTGAAG ATCCTTTTTG ATAATCTCAT	11700
GACCAAAATC CCTTAACGTG AGTTTTTCGT CCACTGAGCG TCAGACCCCG TAGAAAAGAT	11760
CAAAGGATCT TCTTGAGATC CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAAA	11820
ACCACCGCTA CCAGCGGTGG TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTCCGAA	11880
GGTAACTGGC TTCAGCAGAG CGCAGATACC AAATACTGTC CTTCTAGTGT AGCCGTAGTT	11940
AGGCCACCAC TTCAAGAACT CTGTAGCACC GCCTACATAC CTCGCTCTGC TAATCCTGTT	12000
ACCA GTGGCT GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGGTTGGACT CAAGACGATA	12060
GTTACCGGAT AAGGCGCAGC GGTCGGGCTG AACGGGGGGT TCGTGACAC AGCCAGCTT	12120
GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGCGCCAC	12180
GCTTCCCGAA GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA	12240
GCGCACGAGG GAGCTTCCAG GGGGAAACGC CTGGTATCTT TATAGTCCTG TCGGGTTTCG	12300
CCACCTCTGA CTTGAGCGTC GATTTTGTG ATGCTCGTCA GGGGGGCGGA GCCTATGGAA	12360
AAACGCCAGC AACGCGGCCT TTTTACGGTT CCTGGCCTTT TGCTGGCCTT TTGCTCACAT	12420
GTTCTTTCCT GCGTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCCT TTGAGTGAGC	12480
TGATACCGCT CGCCGAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA	12540
AGAGCGCCCA ATACGCAAAC CGCCTCTCCC CGCGCGTTGG CCGATTCATT AATGCAGCTG	12600
GCACGACAGG TTTCCCGACT GGAAAGCGGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA	12660
GCTCACTCAT TAGGCACCCC AGGCTTTACA CTTTATGCTT CCGGCTCGTA TGTTGTGTGG	12720
AATTGTGAGC GGATAACAAT TTCACACAGG AAACAGCTAT GACCATGATT ACGCCAAGCG	12780
CGCAATTAAC CCTCACTAAA GGGAACAAAA GCTG	12814

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15692 base pairs
- (B) TYPE: nucleic acid

- 104 -

(C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = "Targetting vector"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
 (C) INDIVIDUAL ISOLATE: London-FAD APP

(ix) FEATURE:
 (A) NAME/KEY: mat_peptide
 (B) LOCATION: 4807..5151

(ix) FEATURE:
 (A) NAME/KEY: mutation
 (B) LOCATION: replace(4990, "")
 (D) OTHER INFORMATION: /standard_name= "London-FAD"

(ix) FEATURE:
 (A) NAME/KEY: mat_peptide
 (B) LOCATION: 8223..9023

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GAGCTCCACC GCGGTGGCGG CCGCTCTAGA ACTAGTGGAT CCCCCGGGCT GCAGGAATTC	60
TACCGGGGTA GGGGAGGCGC TTTTCCCAAG GCAGTCTGGA GCATGCGCTT TAGCAGCCCC	120
GCTGGCACTT GGCGCTACAC AAGTGGCCTC TGGCCTCGCA CACATTCCAC ATCCACCGGT	180
AGCGCCAACC GGCTCCCTTC TTTGGTGGCC CCTTCGCGCC ACCTTCTACT CCTCCCCTAG	240
TCAGGAAGTT CCCCCCGCC CCGCAGCTCG CGTCGTGCAG GACGTGACAA ATGGAAGTAG	300
CACGTCTCAC TAGTCTCGTG CAGATGGACA GCACCGCTGA GCAATGGAAG CGGGTAGGCC	360
TTTGGGGCAG CGGCCAATAG CAGCTTTGCT CCTTCGCTTT CTGGGCTCAG AGGCTGGGAA	420
GGGGTGGGTC CGGGGGCGGG CTCAGGGGCG GGCTCAGGGG CGGGGCGGGC GCGAAGGTCC	480
TCCGGAGCCC GGCATTCTGC ACGCTTCAAA AGCGCACGTC TGCCGCGCTG TTCTCCTCTT	540
CCTCATCTCC GGGCCTTTTCG ACCTGCAGCG ACCCGCTTAA CAGCGTCAAC AGCGTGCCGC	600
AGATCTTGGT GGCGTGAAAC TCCCGCACCT CTTTGGCAAG CGCCTGTAG AAGCGCGTAT	660
GGCTTCGTAC CCCTGCCATC AACACGCGTC TCGTTCGAC CAGGCTGCGC GTTCTCGCGG	720
CCATAGCAAC CGACGTACGG CGTTGCGCCC TCGCCGGCAG CAAGAAGCCA CGGAAGTCCG	780
CCTGGAGCAG AAAATGCCCA CGCTACTGCG GGTTTATATA GACGGTCCTC ACGGGATGGG	840

- 105 -

GAAAACCACC ACCACGCAAC TGCTGGTGGC CCTGGGTTCG CGCGACGATA TCGTCTACGT	900
ACCCGAGCCG ATGACTTACT GGCAGGTGCT GGGGGCTTCC GAGACAATCG CGAACATCTA	960
CACCACACAA CACCGCCTCG ACCAGGGTGA GATATCGGCC GGGGACGCGG CGGTGGTAAT	1020
GACAAGCGCC CAGATAACAA TGGGCATGCC TTATGCCGTG ACCGACGCCG TTCTGGCTCC	1080
TCATGTCGGG GGGGAGGCTG GGAGTTCACA TGCCCCGCCC CCGGCCCTCA CCCTCATCTT	1140
CGACCGCCAT CCCATCGCCG CCTCCTGTG CTACCCGGCC GCGCGATACC TTATGGGCAG	1200
CATGACCCCC CAGGCCGTGC TGGCGTTCGT GGCCCTCATC CCGCCGACCT TGCCCGGCAC	1260
AAACATCGTG TTGGGGGCCC TTCCGGAGGA CAGACACATC GACCGCCTGG CCAAACGCCA	1320
GCGCCCCGGC GAGCGGCTTG ACCTGGCTAT GCTGGCCGCG ATTCGCCGCG TTTACGGGCT	1380
GCTTGCCAAT ACGGTGCGGT ATCTGCAGGG CGGCGGGTCG TGGTGGGAGG ATTGGGGACA	1440
GCTTTCGGGG ACGGCCGTGC CGCCCCAGGG TGCCGAGCCC CAGAGCAACG CGGGCCCACG	1500
ACCCCATATC GGGGACACGT TATTTACCCT GTTTCGGGCC CCCGAGTTGC TGGCCCCCAA	1560
CGGCGACCTG TATAACGTGT TTGCCTGGGC CTTGGACGTC TTGGCCAAAC GCCTCCGTCC	1620
CATGCACGTC TTTATCCTGG ATTACGACCA ATCGCCCGCC GGCTGCCGGG ACGCCCTGCT	1680
GCAACTTACC TCCGGGATGG TCCAGACCCA CGTCACCACC CCAGGCTCCA TACCGACGAT	1740
CTGCGACCTG GCGCGCACGT TTGCCCGGGA GATGGGGGAG GCTAACTGAA ACACGGAAGG	1800
AGACAATACC GGAAGGAACC CGCGCTATGA CGGCAATAAA AAGACAGAAT AAAACGCACG	1860
GGTGTGGGT CGTTTGTTCA TAAACGCGGG GTTCGGTCCC AGGGCTGGCA CTCTGTGAT	1920
ACCCACCGA GACCCCATTG GGGCCAATAC GCCCGCGTTT CTTCTTTTTC CCCACCCCAA	1980
CCCCAAGTT CGGGTGAAGG CCCAGGGCTC GCAGCCAACG TCGGGGCGGC AAGCCCGCCA	2040
TAGCCACGGG CCCCGTGGGT TAGGGACGGG GTCCCCCATG GGAATGGTT TATGGTTTCGT	2100
GGGGGTATT CTTTGGGCG TTGCGTGGGG TCAGGTCCAC GACTGGACTG AGCAGACAGA	2160
CCCATGGTTT TTGGATGGCC TGGGCATGGA CCGCATGTAC TGGCGCGACA CGAACACCGG	2220
GCGTCTGTGG CTGCCAAACA CCCCCGACCC CAAAAACCA CCGCGCGGAT TTCTGGCGCC	2280
GCCGGACGAA CTAAACCTGA CTACGGCATC TCTGCCCTT CTTGCTGGT ACGAGGAGCG	2340
CTTTTGT TTTTGTATT GTATTGGTCA CCACGGCCGA GTTTCGCGG GACCCCGGCC AGGACCTGCA	2400
GAAATTGATG ATCTATTAAA CAATAAGAT GTCCACTAAA ATGGAAGTTT TTTCTGTCA	2460
TACTTTGTTA AGAAGGGTGA GAACAGAGTA CCTACATTTT GAATGGAAGG ATTGGAGCTA	2520

- 106 -

CGGGGGTGGG	GGTGGGGTGG	GATTAGATAA	ATGCCTGCTC	TTTACTGAAG	GCTCTTTACT	2580
ATTGCTTTAT	GATAATGTTT	CATAGTTGGA	TATCATAATT	TAAACAAGCA	AAACCAAATT	2640
AAGGGCCAGC	TCATTCCCTCC	ACTCATGATC	TATAGATCTA	TAGATCTCTC	GTGGGATCAT	2700
TGTTTTTCTC	TTGATTCCCA	CTTTGTGTTT	TAAGTACTGT	GGTTTCCAAA	TGTGTCAGTT	2760
TCATAGCCTG	AAGAACGAGA	TCAGCAGCCT	CTGTTCCACA	TACACTTCAT	TCTCAGTATT	2820
GTTTTGCCAA	GTTCTAATTC	CATCAGATCA	AGCTTATCGA	TACCGTCGAC	AAGGCGCGCC	2880
ATGTTTAAAC	TTGCGGCCGC	TCTGACCATG	GNNNNNNNNN	NNNNNNNNNA	AGCTTNNNNN	2940
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3000
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3060
NNNNNNNNNN	NNNNNNNNCAT	ATGNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3120
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3180
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3240
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3300
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3360
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3420
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3480
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3540
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNGAATTCT	NNNNNNNNNN	3600
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3660
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3720
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3780
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNGAATT	CNNNNNNNNN	3840
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3900
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3960
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4020
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4080
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNG	AATTCNNNNN	4140
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4200

- 107 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4260
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4320
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4380
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4440
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4500
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4560
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4620
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4680
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4740
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4800
NNNNGTTCTG	GGCTGACAAA	CATCAAGACG	GAAGAGATCT	CTGAAGTGAA	GATGGATGCA	4860
GAATTCCGAC	ATGACTCAGG	ATATGAAGTT	CATCATCAAA	AATTGGTGTT	CTTTGCAGAA	4920
GATGTGGGTT	CAAACAAAGG	TGCAATCATT	GGACTCATGG	TGGGCGGTGT	TGTCATAGCG	4980
ACAGTGATAA	TCATCACCTT	GGTGATGCTG	AAGAAGAAAC	AGTACACATC	CATTCATCAT	5040
GGTGTGGTGG	AGGTTGACGC	CGCTGTCACC	CCAGAGGAGC	GCCACCTGTC	CAAGATGCAG	5100
CAGAACGGCT	ACGAAAATCC	AACCTACAAG	TTCTTTGAGC	AGATGCAGAA	CTAGACCCCC	5160
GCCACAGCAG	CCTCTGAAGT	TGGACAGCAA	AACCATTGCT	TCACTACCCA	TCGGTGTCCA	5220
TTTATAGAAT	AATGTGGGAA	GAAACAAACC	CGTTTTATGA	TTTACTCATT	ATCGCCTTTT	5280
GACAGCTGTG	CTGTAACACA	AGTAGATGCC	TGAACTTGAA	TTAATCCACA	CATCAGTAAT	5340
GTATTCTATC	TCTCTTTACA	TTTTGGTCTC	TATACTACAT	TATTAATGGG	TTTTGTGTAC	5400
TGTAAAGAAT	TTAGCTGTAT	CAAAC TAGTG	CATGAATAGA	TTCTCTCCTG	ATTATTTATC	5460
ACATAGCCCC	TTAGCCAGTT	GTATATTATT	CTTGTGGTTT	GTGACCCAAT	TAAGTCCTAC	5520
TTTACATATG	CTTTAAGAAT	CGATGGGGGA	TGCTTCATGT	GAACGTGGGA	GTTTCAGCTGC	5580
TTCTCTTGCC	TAAGTATTCC	TTTCCTGATC	ACTATGCATT	TTAAAGTTAA	ACATTTTTTAA	5640
GTATTTTACA	TGCTTTAGAG	AGATTTTTTT	TCCATGACTG	CATTTTACTG	TACAGATTGC	5700
TGCTTCTGCT	ATATTTGTGA	TATAGGAATT	AAGAGGATAC	ACACGTTTGT	TTCTTCGTGC	5760
CTGTTTTATG	TGCACACATT	AGGCATTGAG	ACTTCAAGCT	TTTCTTTTTT	TGTCCACGTA	5820
TCTTTGGGTC	TTTGATAAAG	AAAAGAATCC	CTGTTTATTG	TAAGCACTTT	TACGGGGCGG	5880

- 108 -

GTGGGGAGGG GTGCTCTGCT GGTCTTCAAT TACCAAGAAT TCTCCAAAAC AATTTTCTGC	5940
AGGATGATTG TACAGAATCA TTGCTTATGA CATGATCGCT TTCTACACTG TATTACATAA	6000
ATAAATTAAA TAAATAAACC CCGGGCAAGA CTTTTCTTTG AAGGATGACT ACAGACATTA	6060
AATAATCGAA GTAATTTTGG GTGGGGAGAA GAGGCAGATT CAATTTTCTT TAACCAGTCT	6120
GAAGTTTCAT TTATGATACA AAAGAAGATG AAAATGGAAG TGGCAATATA AGGGGATGAG	6180
GAAGGCATGC CTGGACAAAC CCTTCTTTTA AGATGTGTCT TCAATTTGTA TAAAATGGTG	6240
TTTTCATGTA AATAAATACA TTCTTGGAGG AGCCACATTG TGCTGGTGTG AATGATTCCA	6300
TAGTAACAAT CTTGACCATT TACTGACGTA CAGACCAGTG AGAAGTCTTC GCATGTTGGG	6360
TACCCACACC TGTGTGTCT TAATTGCAAG TCTGAGTAGG AAGTTGGGGC CAACATGTGT	6420
CTCCAGTGC TGGGAAAATA TTTCATAGAC CTAATTTACA GTCTTTACTT GATCTAAAAC	6480
ATTTTGCTGC CATATTTTGG CCCTCAAGTT TGTCCCAAAT GAGAGACAAA GGGAAAAGTT	6540
CCAGGGAAAT AAAAATTAAG ACAGCTGATT ATCTGTAAAG CATGGTTTCT CATCCTGAAC	6600
GCTACTAACA TTTTGCAGGG AATAATTCCT TGTGAAGGG AGTTGTCCTG ACCAGTGTAG	6660
GATATTTATT TATTTTATTT ATGTTTTTTG AGACGGAGTC TCGCTCTGTC ACCCAGGCTG	6720
GAGTGCAGTG GCACAATCTC GGCTCACTGC AAGCTCCGCC TCCCGGGTTC ACGCCATTCT	6780
CCTGCCTCAG CCTCCTGAAT AGCTGGGACT CTAGGTGCCC GCCACCACGC CCGGCTAATT	6840
TTTTGTATTT TTAGTAGAGA CGGGGTTTCA CCGTGTTAGC CAGGACAGTC TTGGTCTCCT	6900
GACCTCGTGA TCTGCCTGCC TCGGCCTCCC AAAGTGCTGA GATTACAGGC GTGCAAGCCG	6960
CGCCCAGCCA GTGCTCTCCT TTAAAAGTA GCCCATTGGC TGGGCGCAGT GGCTCACGCC	7020
TGTAATCCCA GCACTTTGGG AGGCTGAGGC GGGTGGATCA CGAGGTCAGG AGATCAAGAA	7080
TATCCTGGCC AATATGGTGA AACCCCATCT CTAATAAAAA TACAAAAAAA AAAAAAAAAA	7140
AAAAAAGGCC GGGCATGGTG GCGGGCGCTT GTAGTCCCAG CTAATCAGGA GGCTGAGGCA	7200
GGAGAATGGT GTGCACCTGG GAGGCGGAGG TTGCAGTGAG CTGAGATCGC GCCACTGCAC	7260
TCCAGCCTGG GAGACAGAGC GAGACTCCGT CTCAATAAAT AAATAAATAA ATAAATAAAA	7320
GGAGGGCCTG GCACGAATGA CATGCAGGGA AGGCAGTGAG CAGGTGGAGG TCCCTGTACT	7380
CGTTGTGGTG CCTTATCTAC CAGGCGGTTG AGTTGACGTC TTTGTGGACA GAATTCGAGC	7440
TCGGTACCCG GGGATCCTCT AGAGTCGACC TTAAGGTCGA CGGTATCGAT AAGCTTGGGC	7500
TTGAACATCG AGCGCCAGGG CTCCGTAAAG CTACTAGAGC ACAGGCGGTG CCCCACGTC	7560

CTGGGGCCTC	TCCACTAATA	ACGGCTACTT	CCAATTGATT	GGACGCGCCA	TCTTGCCTGC	7620
CTTATGCATA	TTCAGCGGTG	AACTGAATAT	TCATGAACGA	GGCCCGTCCC	GTCCCTCCCT	7680
CCTTCCCCCC	ACCCCCGGAA	CCCCTCCGG	AGGACCCGAA	GGGCCCCGCC	TTCATTACCG	7740
ATGCGTAGGA	CAAACCATTT	TCCCGATGTG	TGTGGGGGGA	TACTAATGAG	AGACTTTAGC	7800
TGAAAAATGA	GCCTGAACTC	CGAAGCTGAG	TAAAAATGGC	CTAACTTTAT	CCTCCGTTCT	7860
GTAAGTCCTC	GGTTTGAGTG	CACGGGAAAC	CCGAAAGGAG	GACGACAGGA	CCAGGACATT	7920
CTCCTCCTCC	TGTCGCGTCA	GAAAGAACAC	CCAACCAGGG	AGCCGGAGCC	CTAGCGTCAA	7980
CAACTCCGCC	GCGCGCGCTC	CGTGTAGGCC	GGTGCGGGCG	GCCCCGTAGC	GCAAGGGAGG	8040
GCGGGAAAGG	AAGGGGCGGG	ACACAAGGGC	GAATCTATAA	AGGGCGTCAC	TCAGCCAGTT	8100
CTCTCCTCAG	AAGCGCCGAG	AGCGCGACCG	GGACGGTTGG	AGAAGAAGGT	GGCTCCCGGA	8160
AGGGGGAGAG	ACAAACTGCC	GTAACCTCTG	CCGTTTCAAG	TCATCGAATT	CCTGCAGCCA	8220
ATATGGGATC	GGCCATTGAA	CAAGATGGAT	TGCACGCAGG	TTCTCCGGCC	GCTTGGGTGG	8280
AGAGGCTATT	CGGCTATGAC	TGGGCACAAC	AGACAATCGG	CTGCTCTGAT	GCCGCCGTGT	8340
TCCGGCTGTC	AGCGCAGGGG	CGCCCGGTTT	TTTTTGTCAA	GACCGACCTG	TCCGGTGCCC	8400
TGAATGAACT	GCAGGACGAG	GCAGCGCGGC	TATCGTGGCT	GGCCACGACG	GGCGTTTCTT	8460
GCGCAGCTGT	GCTCGACGTT	GTCAGTGAAG	CGGGAAGGGA	CTGGCTGCTA	TTGGGCGAAG	8520
TGCCGGGGCA	GGATCTCCTG	TCATCTCACC	TTGCTCCTGC	CGAGAAAGTA	TCCATCATGG	8580
CTGATGCAAT	GCGGCGGCTG	CATACGCTTG	ATCCGGCTAC	CTGCCCATTG	GACCACCAAG	8640
CGAAACATCG	CATCGAGCGA	GCACGTACTC	GGATGGAAGC	CGGTCTTGTC	GATCAGGATG	8700
ATCTGGACGA	AGAGCATCAG	GGGCTCGCGC	CAGCCGAACT	GTTGCGCCAGG	CTCAAGGCGC	8760
GCATGCCCCA	CGGCGAGGAT	CTCGTCGTGA	CCCATGGCGA	TGCCTGCTTG	CCGAATATCA	8820
TGGTGGAAAA	TGGCCGCTTT	TCTGGATTCA	TCGACTGTGG	CCGGCTGGGT	GTGGCGGACC	8880
GCTATCAGGA	CATAGCGTTG	GCTACCCGTG	ATATTGCTGA	AGAGCTTGGC	GGCGAATGGG	8940
CTGACCGCTT	CCTCGTGCTT	TACGGTATCG	CCGCTCCCGA	TTCGACGCGC	ATCGCCTTCT	9000
ATCGCCTTCT	TGACGAGTTC	TTCTGAGGGG	ATCAATTCTC	TAGAGCTCGC	TGATCAGCCT	9060
CGACTGTGCC	TTCTAGTTGC	CAGCCATCTG	TTGTTTGCCC	CTCCCCCGTG	CCTTCCTTGA	9120
CCCTGGAAGG	TGCCACTCCC	ACTGTCCTTT	CCTAATAAAA	TGAGGAAATT	GCATCGCATT	9180
GTCTGAGTAG	GTGTCATTCT	ATTCTGGGGG	GTGGGGTGGG	GCAGGACAGC	AAGGGGGAGG	9240

- 110 -

ATTGGGAAGA	CAATAGCAGG	CATGCTGGGG	ATGCGGTGGG	CTCTATGGCT	TCTGAGGCGG	9300
AAAGAACCAG	CTGGGGCTCG	AGAGATCTTC	ACAANGATAG	GAAGGAGAGG	AAGTGGGGCT	9360
CTGTTGATAG	TTCTTGCTGA	GCAGAAGCCN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9420
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9480
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9540
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9600
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9660
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9720
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9780
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9840
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9900
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9960
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10020
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10080
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10140
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10200
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10260
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10320
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10380
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10440
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10500
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10560
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10620
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10680
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNGG	10740
ATCCNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10800
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10860
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10920

[illegible]

- 112 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12660
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12720
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNC	CATGGTCTAG	AACTAGTGGA	TCCCCGGGGC	12780
TGCAGGAATT	CGATATCAAG	CTTATCGATA	CCGTCGACCT	CGAGGGGGGG	CCCGGTACCC	12840
AATTGCCCCCT	ATAGTGAGTC	GTATTACGCG	CGCTCACTGG	CCGTCGTTTT	ACAACGTCGT	12900
GACTGGGAAA	ACCCTGGCGT	TACCCAACCT	AATCGCCTTG	CAGCACATCC	CCCTTTCGCC	12960
AGCTGGCGTA	ATAGCGAAGA	GGCCCGCACC	GATCGCCCTT	CCCAACAGTT	GCGCAGCCTG	13020
AATGGCGAAT	GGGACGCGCC	CTGTAGCGGC	GCATTAAGCG	CGGCGGGTGT	GGTGGTTACG	13080
CGCAGCGTGA	CCGCTACACT	TGCCAGCGCC	CTAGCGCCCG	CTCCTTTCGC	TTTCTTCCCT	13140
TCCTTTCTCG	CCACGTTTCGC	CGGCTTTCCC	CGTCAAGCTC	TAAATCGGGG	GCTCCCTTTA	13200
GGGTTCGGAT	TTAGTGCTTT	ACGGCACCTC	GACCCCAAAA	AACTTGATTA	GGGTGATGGT	13260
TCACGTAGTG	GGCCATCGCC	CTGATAGACG	GTTTTTCGCC	CTTGACGTT	GGAGTCCACG	13320
TTCTTTAATA	GTGGACTCTT	GTTCCAAACT	GGAACAACAC	TCAACCCTAT	CTCGGTCTAT	13380
TCTTTTGATT	TATAAGGGAT	TTTGCCGATT	TCGGCCTATT	GGTTAAAAA	TGAGCTGATT	13440
TAACAAAAAT	TTAACGCGAA	TTTTAACAAA	ATATTAACGC	TTACAATTTA	GGTGGCACTT	13500
TTCGGGGAAA	TGTGCGCGGA	ACCCCTATTT	GTTTATTTTT	CTAAATACAT	TCAAATATGT	13560
ATCCGCTCAT	GAGACAATAA	CCCTGATAAA	TGCTTCAATA	ATATTGAAAA	AGGAAGAGTA	13620
TGAGTATTCA	ACATTTCCGT	GTCGCCCTTA	TTCCCTTTTT	TGCGGCATTT	TGCCTTCCTG	13680
TTTTTGCTCA	CCCAGAAACG	CTGGTGAAAG	TAAAAGATGC	TGAAGATCAG	TTGGGTGCAC	13740
GAGTGGGTTA	CATCGAACTG	GATCTCAACA	GCGGTAAGAT	CCTTGAGAGT	TTTCGCCCCG	13800
AAGAACGTTT	TCCAATGATG	AGCACTTTTA	AAGTTCTGCT	ATGTGGCGCG	GTATTATCCC	13860
GTATTGACGC	CGGGCAAGAG	CAACTCGGTC	GCCGCATACA	CTATTCTCAG	AATGACTTGG	13920
TTGAGTACTC	ACCAGTCACA	GAAAAGCATC	TTACGGATGG	CATGACAGTA	AGAGAATTAT	13980
GCAGTGCTGC	CATAACCATG	AGTGATAACA	CTGCGGCCAA	CTTACTTCTG	ACAACGATCG	14040
GAGGACCGAA	GGAGCTAACC	GCTTTTTTTC	ACAACATGGG	GGATCATGTA	ACTCGCCTTG	14100
ATCGTTGGGA	ACCGGAGCTG	AATGAAGCCA	TACCAAACGA	CGAGCGTGAC	ACCACGATGC	14160
CTGTAGCAAT	GGCAACAACG	TTGCGCAAAC	TATTAACCTG	CGAACTACTT	ACTCTAGCTT	14220
CCCCGGCAACA	ATTAATAGAC	TGGATGGAGG	CGGATAAAGT	TGCAGGACCA	CTTCTGCGCT	14280

- 113 -

CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC	14340
GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA GTTATCTACA	14400
CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG ATAGGTGCCT	14460
CACTGATTAA GCATTGGTAA CTGTCAGACC AAGTTTACTC ATATATACTT TAGATTGATT	14520
TAAAACCTCA TTTTAAATTT AAAAGGATCT AGGTGAAGAT CCTTTTGTAT AATCTCATGA	14580
CCAAAATCCC TTAACGTGAG TTTTCGTTCC ACTGAGCGTC AGACCCCGTA GAAAAGATCA	14640
AAGGATCTTC TTGAGATCCT TTTTCTCTGC GCGTAATCTG CTGCTTGCAA ACAAAAAAAC	14700
CACCGCTACC AGCGGTGGTT TGTTTGCCGG ATCAAGAGCT ACCAACTCTT TTTCCGAAGG	14760
TAACTGGCTT CAGCAGAGCG CAGATACCAA ATACTGTCCT TCTAGTGTAG CCGTAGTTAG	14820
GCCACCACTT CAAGAACTCT GTAGCACCGC CTACATACCT CGCTCTGCTA ATCCTGTTAC	14880
CAGTGGCTGC TGCCAGTGGC GATAAGTCGT GTCTTACCGG GTTGGACTCA AGACGATAGT	14940
TACCGGATAA GGCAGCGCG TCGGGCTGAA CGGGGGGTTC GTGCACACAG CCCAGCTTGG	15000
AGCGAACGAC CTACACCGAA CTGAGATACC TACAGCGTGA GCTATGAGAA AGCGCCACGC	15060
TTCCCGAAGG GAGAAAGGCG GACAGGTATC CGGTAAGCGG CAGGGTCGGA ACAGGAGAGC	15120
GCACGAGGGA GCTTCCAGGG GGAAACGCCT GGTATCTTTA TAGTCCTGTC GGGTTTCGCC	15180
ACCTCTGACT TGAGCGTCGA TTTTGTGAT GCTCGTCAGG GGGGCGGAGC CTATGGAAAA	15240
ACGCCAGCAA CGCGGCCTTT TTACGGTTCC TGGCCTTTTG CTGGCCTTTT GCTCACATGT	15300
TCTTTCCTGC GTTATCCCCT GATTCTGTGG ATAACCGTAT TACCGCCTTT GAGTGAGCTG	15360
ATACCGCTCG CCGCAGCCGA ACGACCGAGC GCAGCGAGTC AGTGAGCGAG GAAGCGGAAG	15420
AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCAATTA TGCAGCTGGC	15480
ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC	15540
TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA	15600
TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCGCG	15660
CAATTAACCC TCACTAAAGG GAACAAAAGC TG	15692

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15692 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- 114 -

(ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = "Targetting vector"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
 (C) INDIVIDUAL ISOLATE: Swedish/London-FAD APP

(ix) FEATURE:
 (A) NAME/KEY: mat_peptide
 (B) LOCATION: 4807..5151

(ix) FEATURE:
 (A) NAME/KEY: mutation
 (B) LOCATION: replace(4849, "")
 (D) OTHER INFORMATION: /standard_name= "Swedish-FAD"

(ix) FEATURE:
 (A) NAME/KEY: mutation
 (B) LOCATION: replace(4989, "")
 (D) OTHER INFORMATION: /standard_name= "London-FAD"

(ix) FEATURE:
 (A) NAME/KEY: mat_peptide
 (B) LOCATION: 8223..9023

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GAGCTCCACC GCGGTGGCGG CCGCTCTAGA ACTAGTGGAT CCCCCGGGCT GCAGGAATTC	60
TACCGGGGTA GGGGAGGCGC TTTTCCCAAG GCAGTCTGGA GCATGCGCTT TAGCAGCCCC	120
GCTGGCACTT GCGCTACAC AAGTGGCCTC TGGCCTCGCA CACATTCCAC ATCCACCGGT	180
AGCGCCAACC GGCTCCCTTC TTTGGTGGCC CCTTCGCGCC ACCTTCTACT CCTCCCCTAG	240
TCAGGAAGTT CCCCCCGCC CCGCAGCTCG CGTCGTGCAG GACGTGACAA ATGGAAGTAG	300
CACGTCTCAC TAGTCTCGTG CAGATGGACA GCACCGCTGA GCAATGGAAG CGGGTAGGCC	360
TTTGGGGCAG CGGCAATAG CAGCTTTGCT CCTTCGCTTT CTGGGCTCAG AGGCTGGGAA	420
GGGGTGGGTC CGGGGGCGGG CTCAGGGGCG GGCTCAGGGG CGGGGCGGGC GCGAAGGTCC	480
TCCGGAGCCC GGCATTCTGC ACGCTTCAAA AGCGCACGTC TGCCGCGCTG TTCTCCTCTT	540
CCTCATCTCC GGGCCTTTTCG ACCTGCAGCG ACCCGCTTAA CAGCGTCAAC AGCGTGCCGC	600
AGATCTTGGT GCGGTGAAAC TCCCGCACCT CTTTGGCAAG CGCCTTGTAG AAGCGCGTAT	660
GGCTTCGTAC CCCTGCCATC AACACGCGTC TCGGTTTCGAC CAGGCTGCGC GTTCTCGCGG	720
CCATAGCAAC CGACGTACGG CGTTGCGCCC TCGCCGGCAG CAAGAAGCCA CGGAAGTCCG	780

- 115 -

CCTGGAGCAG	AAAATGCCCA	CGCTACTGCG	GGTTTATATA	GACGGTCCTC	ACGGGATGGG	840
GAAAACCACC	ACCACGCAAC	TGCTGGTGGC	CCTGGGTTTCG	CGCGACGATA	TCGTCTACGT	900
ACCCGAGCCG	ATGACTTACT	GGCAGGTGCT	GGGGGCTTCC	GAGACAATCG	CGAACATCTA	960
CACCACACAA	CACCGCCTCG	ACCAGGGTGA	GATATCGGCC	GGGGACGCGG	CGGTGGTAAT	1020
GACAAGCGCC	CAGATAACAA	TGGGCATGCC	TTATGCCGTG	ACCGACGCCG	TTCTGGCTCC	1080
TCATGTCGGG	GGGGAGGCTG	GGAGTTCACA	TGCCCCGCCC	CCGGCCCTCA	CCCTCATCTT	1140
CGACCGCCAT	CCCATCGCCG	CCCTCCTGTG	CTACCCGGCC	GCGCGATACC	TTATGGGCAG	1200
CATGACCCCC	CAGGCCGTGC	TGGCGTTTCGT	GGCCCTCATC	CCGCCGACCT	TGCCCCGGCAC	1260
AAACATCGTG	TTGGGGGGCCC	TTCCGGAGGA	CAGACACATC	GACCGCCTGG	CCAAACGCCA	1320
GCGCCCCGGC	GAGCGGCTTG	ACCTGGCTAT	GCTGGCCGCG	ATTCGCCGCG	TTTACGGGCT	1380
GCTTGCCAAT	ACGGTGCGGT	ATCTGCAGGG	CGGCGGGTCG	TGGTGGGAGG	ATTGGGGACA	1440
GCTTTCGGGG	ACGGCCGTGC	CGCCCCAGGG	TGCCGAGCCC	CAGAGCAACG	CGGGCCCACG	1500
ACCCCATATC	GGGGACACGT	TATTTACCCT	GTTTCGGGGC	CCCGAGTTGC	TGGCCCCCAA	1560
CGGCGACCTG	TATAACGTGT	TTGCCTGGGC	CTTGACGTC	TTGGCCAAAC	GCCTCCGTCC	1620
CATGCACGTC	TTTATCCTGG	ATTACGACCA	ATCGCCCGCC	GGCTGCCGGG	ACGCCCTGCT	1680
GCAACTTACC	TCCGGGATGG	TCCAGACCCA	CGTCACCACC	CCAGGCTCCA	TACCGACGAT	1740
CTGCGACCTG	GCGCGCACGT	TTGCCCGGGA	GATGGGGGAG	GCTAACTGAA	ACACGGAAGG	1800
AGACAATACC	GGAAGGAACC	CGCGCTATGA	CGGCAATAAA	AAGACAGAAT	AAAACGCACG	1860
GGTGTGGGT	CGTTTGTTCA	TAAACGCGGG	GTTTCGGTCCC	AGGGCTGGCA	CTCTGTCGAT	1920
ACCCCAACCGA	GACCCCATTG	GGGCCAATAC	GCCCCGCTTT	CTTCCTTTTC	CCCACCCCAA	1980
CCCCCAAGTT	CGGGTGAAGG	CCCAGGGCTC	GCAGCCAACG	TCGGGGCGGC	AAGCCCCGCA	2040
TAGCCACGGG	CCCCGTGGGT	TAGGGACGGG	GTCCCCCATG	GGGAATGGTT	TATGGTTTCGT	2100
GGGGGTATT	CTTTTGGGCG	TTGCGTGGGG	TCAGGTCCAC	GACTGGACTG	AGCAGACAGA	2160
CCCATGGTTT	TTGGATGGCC	TGGGCATGGA	CCGCATGTAC	TGGCGCGACA	CGAACACCGG	2220
GCGTCTGTGG	CTGCCAAACA	CCCCCGACCC	CAAAAAACCA	CCGCGCGGAT	TTCTGGCGCC	2280
GCCGGACGAA	CTAAACCTGA	CTACGGCATC	TCTGCCCCCTT	CTTCGCTGGT	ACGAGGAGCG	2340
CTTTTGTTTT	GTATTGGTCA	CCACGGCCGA	GTTTCCGCGG	GACCCCGGCC	AGGACCTGCA	2400
GAAATTGATG	ATCTATTAAA	CAATAAAGAT	GTCCACTAAA	ATGGAAGTTT	TTTCCTGTCA	2460

- 116 -

TACTTTGTTA AGAAGGGTGA GAACAGAGTA CCTACATTTT GAATGGAAGG ATTGGAGCTA	2520
CGGGGGTGGG GGTGGGGTGG GATTAGATAA ATGCCTGCTC TTTACTGAAG GCTCTTTACT	2580
ATTGCTTTAT GATAATGTTT CATAGTTGGA TATCATAATT TAAACAAGCA AAACCAAATT	2640
AAGGGCCAGC TCATTCTCC ACTCATGATC TATAGATCTA TAGATCTCTC GTGGGATCAT	2700
TGTTTTTCTC TTGATTCCCA CTTTGTGTTC TAAGTACTGT GGTTCCTCAA TGTGTCAGTT	2760
TCATAGCCTG AAGAACGAGA TCAGCAGCCT CTGTTCCACA TACACTTCAT TCTCAGTATT	2820
GTTTTGCCAA GTTCTAATTC CATCAGATCA AGCTTATCGA TACCGTCGAC AAGGCGCGCC	2880
ATGTTTAAAC TTGCGGCCGC TCTGACCATG GNNNNNNNNN NNNNNNNNNN AGCTTNNNNN	2940
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3000
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3060
NNNNNNNNNN NNNNNNNNNN ATGNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3120
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3180
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3240
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3300
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3360
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3420
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3480
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3540
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3600
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3660
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3720
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3780
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3840
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3900
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	3960
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	4020
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	4080
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN AATTCNNNNN	4140

- 117 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4200
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4260
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4320
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4380
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4440
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4500
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4560
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4620
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4680
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4740
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4800
NNNNGTTCTG	GGCTGACAAA	CATCAAGACG	GAAGAGATCT	CTGAAGTGAA	TCTAGATGCA	4860
GAATTCCGAC	ATGACTCAGG	ATATGAAGTT	CATCATCAAA	AATTGGTGT	CTTTGCAGAA	4920
GATGTGGGTT	CAAACAAAGG	TGCAATCATT	GGACTCATGG	TGGGCGGTGT	TGTCATAGCG	4980
ACAGTGATAA	TCATCACCTT	GGTGATGCTG	AAGAAGAAAC	AGTACACATC	CATTTCATCAT	5040
GGTGTGGTGG	AGGTTGACGC	CGCTGTCACC	CCAGAGGAGC	GCCACCTGTC	CAAGATGCAG	5100
CAGAACGGCT	ACGAAAATCC	AACCTACAAG	TTCTTTGAGC	AGATGCAGAA	CTAGACCCCC	5160
GCCACAGCAG	CCTCTGAAGT	TGGACAGCAA	AACCATTGCT	TCACTACCCA	TCGGTGTCCA	5220
TTTATAGAAT	AATGTGGGAA	GAAACAAACC	CGTTTTATGA	TTTACTCATT	ATCGCCTTTT	5280
GACAGCTGTG	CTGTAACACA	AGTAGATGCC	TGAACTTGAA	TTAATCCACA	CATCAGTAAT	5340
GTATTCTATC	TCTCTTTACA	TTTTGGTCTC	TATACTACAT	TATTAATGGG	TTTTGTGTAC	5400
TGTAAAGAAT	TTAGCTGTAT	CAAAC TAGTG	CATGAATAGA	TTCTCTCCTG	ATTATTTATC	5460
ACATAGCCCC	TTAGCCAGTT	GTATATTATT	CTTGTGGTTT	GTGACCCAAT	TAAGTCCTAC	5520
TTTACATATG	CTTTAAGAAT	CGATGGGGGA	TGCTTCATGT	GAACGTGGGA	GTTCAGCTGC	5580
TTCTCTTGCC	TAAGTATTCC	TTTCCTGATC	ACTATGCATT	TTAAAGTTAA	ACATTTTTAA	5640
GTATTTTACA	TGCTTTAGAG	AGATTTTTTT	TCCATGACTG	CATTTTACTG	TACAGATTGC	5700
TGCTTCTGCT	ATATTTGTGA	TATAGGAATT	AAGAGGATAC	ACACGTTTGT	TTCTTCGTGC	5760
CTGTTTTATG	TGCACACATT	AGGCATTGAG	ACTTCAAGCT	TTTCTTTTTT	TGTCCACGTA	5820

- 118 -

TCTTTGGGTC	TTTGATAAAG	AAAAGAATCC	CTGTTTCATTG	TAAGCACTTT	TACGGGGCGG	5880
GTGGGGAGGG	GTGCTCTGCT	GGTCTTCAAT	TACCAAGAAT	TCTCCAAAAC	AATTTTCTGC	5940
AGGATGATTG	TACAGAATCA	TTGCTTATGA	CATGATCGCT	TTCTACACTG	TATTACATAA	6000
ATAAATTAAA	TAAAATAACC	CCGGGCAAGA	CTTTTCTTTG	AAGGATGACT	ACAGACATTA	6060
AATAATCGAA	GTAATTTTGG	GTGGGGAGAA	GAGGCAGATT	CAATTTTCTT	TAACCAGTCT	6120
GAAGTTTCAT	TTATGATACA	AAAGAAGATG	AAAATGGAAG	TGGCAATATA	AGGGGATGAG	6180
GAAGGCATGC	CTGGACAAAC	CCTTCTTTTA	AGATGTGTCT	TCAATTTGTA	TAAAATGGTG	6240
TTTTTCATGTA	AATAAATACA	TTCTTGGAGG	AGCCACATTG	TGCTGGTGTG	AATGATTCCA	6300
TAGTAACAAT	CTTGACCATT	TACTGACGTA	CAGACCAGTG	AGAAGTCTTC	GCATGTTGGG	6360
TACCCACACC	TGTTGTGTCT	TAATTGCAAG	TCTGAGTAGG	AAGTTGGGGC	CAACATGTGT	6420
CTCCCAGTGC	TGGGAAAATA	TTTCATAGAC	CTAATTTACA	GTCTTTACTT	GATCTAAAAC	6480
ATTTTGCTGC	CATATTTTGG	CCCTCAAGTT	TGTCCCAAAT	GAGAGACAAA	GGGAAAAGTT	6540
CCAGGGAAAT	AAAAATTAAG	ACAGCTGATT	ATCTGTAAAG	CATGGTTTCT	CATCCTGAAC	6600
GCTACTAACA	TTTTGCAGGG	AATAATTCCT	TGTTGAAGGG	AGTTGTCCTG	ACCAGTGTAG	6660
GATATTTATT	TATTTTATTT	ATGTTTTTTG	AGACGGAGTC	TCGCTCTGTC	ACCCAGGCTG	6720
GAGTGCACTG	GCACAATCTC	GGCTCACTGC	AAGCTCCGCC	TCCCGGGTTC	ACGCCATTCT	6780
CCTGCCTCAG	CCTCCTGAAT	AGCTGGGACT	CTAGGTGCCC	GCCACCACGC	CCGGCTAATT	6840
TTTTGTATTT	TTAGTAGAGA	CGGGGTTTCA	CCGTGTTAGC	CAGGACAGTC	TTGGTCTCCT	6900
GACCTCGTGA	TCTGCCTGCC	TCGGCCTCCC	AAAGTGCTGA	GATTACAGGC	GTGCAAGCCG	6960
CGCCCAGCCA	GTGCTCTCCT	TTTAAAAGTA	GCCCATTGGC	TGGGCGCAGT	GGCTCACGCC	7020
TGTAATCCCA	GCACTTTGGG	AGGCTGAGGC	GGGTGGATCA	CGAGGTCAGG	AGATCAAGAA	7080
TATCCTGGCC	AATATGGTGA	AACCCCATCT	CTACTAAAAA	TACAAAAAAA	AAAAAAAAAA	7140
AAAAAAGGCC	GGGCATGGTG	GCGGGCGCTT	GTAGTCCCAG	CTACTCAGGA	GGCTGAGGCA	7200
GGAGAATGGT	GTGCACCTGG	GAGGCGGAGG	TTGCAGTGAG	CTGAGATCGC	GCCACTGCAC	7260
TCCAGCCTGG	GAGACAGAGC	GAGACTCCGT	CTCAATAAAT	AAATAAATAA	ATAAATAAAA	7320
GGAGGGCCTG	GCACGAATGA	CATGCAGGGA	AGGCAGTGAG	CAGGTGGAGG	TCCCTGTACT	7380
CGTTGTGGTG	CCTTATCTAC	CAGGCGGTTG	AGTTGACGTC	TTTGTGGACA	GAATTCGAGC	7440
TCGGTACCCG	GGGATCCTCT	AGAGTCGACC	TTAAGGTCGA	CGGTATCGAT	AAGCTTGGGC	7500

- 119 -

TTGAACATCG AGCGCCAGGG CTCCGTAAAG CTACTAGAGC ACAGGCGGTG CCCC AACGTC	7560
CTGGGGCCTC TCCACTAATA ACGGCTACTT CCAATTGATT GGACGCGCCA TCTTGCTGC	7620
CTTATGCATA TTCAGCGGTG AACTGAATAT TCATGAACGA GGCCCGTCCC GTCCCTCCCT	7680
CCTTCCCCC ACCCCCGGAA CCCGCTCCGG AGGACCCGAA GGGCCCCGCC TTCATTACCG	7740
ATGCGTAGGA CAAACCATT TCCCGATGTG TGTGGGGGGA TACTAATGAG AGACTTTAGC	7800
TGAAAAATGA GCCTGAACTC CGAAGCTGAG TAAAAATGGC CTAAC'TTTAT CCTCCGTTCT	7860
GTAAGTCCTC GGT'TTGAGTG CACGGGAAAC CCGAAAGGAG GACGACAGGA CCAGGACATT	7920
CTCCTCCTCC TGTCGCGTCA GAAAGAACAC CCAACCAGGG AGCCGGAGCC CTAGCGTCAA	7980
CAACTCCGCC GCGCGCGCTC CGTG'TAGGCC GGTGCGGGCG GCCCCGTAGC GCAAGGGAGG	8040
GCGGGAAAGG AAGGGGCGGG ACACAAGGGC GAATCTATAA AGGGCGTCAC TCAGCCAGTT	8100
CTCTCCTCAG AAGCGCCGAG AGCGCGACCG GGACGGTTGG AGAAGAAGGT GGCTCCCGGA	8160
AGGGGGAGAG ACAAACTGCC GTAACCTCTG CCGTTCAGGA TCATCGAATT CCTGCAGCCA	8220
ATATGGGATC GGCCATTGAA CAAGATGGAT TGCACGCAGG TTCTCCGGCC GCTTGGGTGG	8280
AGAGGCTATT CGGCTATGAC TGGGCACAAC AGACAATCGG CTGCTCTGAT GCCGCCGTGT	8340
TCCGGCTGTC AGCGCAGGGG CGCCCGGTTT TTTTGTCAA GACCGACCTG TCCGGTGCCC	8400
TGAATGAACT GCAGGACGAG GCAGCGCGGC TATCGTGGCT GGCCACGACG GCGGTTCTT	8460
GCGCAGCTGT GCTCGACGTT GTCACTGAAG CGGGAAGGGA CTGGCTGCTA TTGGGCGAAG	8520
TGCCGGGGCA GGATCTCCTG TCATCTCACC TTGCTCCTGC CGAGAAAGTA TCCATCATGG	8580
CTGATGCAAT GCGGCGGCTG CACACGCTTG ATCCGGCTAC CTGCCCATTG GACCACCAAG	8640
CGAAACATCG CATCGAGCGA GCACGTACTC GGATGGAAGC CGGTCTTGTC GATCAGGATG	8700
ATCTGGACGA AGAGCATCAG GGGCTCGCGC CAGCCGAACT GTTCGCCAGG CTCAAGGCGC	8760
GATGCCCCGA CGGCGAGGAT CTCGTCGTGA CCCATGGCGA TGCCTGCTTG CCGAATATCA	8820
TGGTGGAAAA TGGCCGCTTT TCTGGATTCA TCGACTGTGG CCGGCTGGGT GTGGCGGACC	8880
GCTATCAGGA CATAGCGTTG GCTACCCGTG ATATTGCTGA AGAGCTTGGC GGCGAATGGG	8940
CTGACCGCTT CCTCGTGCTT TACGGTATCG CCGCTCCCGA TTCGCAGCGC ATCGCCTTCT	9000
ATCGCCTTCT TGACGAGTTC TTCTGAGGGG ATCAATTCTC TAGAGCTCGC TGATCAGCCT	9060
CGACTGTGCC TTCTAGTTGC CAGCCATCTG TTGTTTGCCC CTCCCCGTG CCTTCCTTGA	9120
CCCTGGAAGG TGCCACTCCC ACTGTCCTTT CCTAATAAAA TGAGGAAATT GCATCGCATT	9180

- 120 -

GTCTGAGTAG	GTGTCATTCT	ATTCTGGGGG	GTGGGGTGGG	GCAGGACAGC	AAGGGGGAGG	9240
ATTGGGAAGA	CAATAGCAGG	CATGCTGGGG	ATGCGGTGGG	CTCTATGGCT	TCTGAGGCGG	9300
AAAGAACCAG	CTGGGGCTCG	AGAGATCTTC	ACAANGATAG	GAAGGAGAGG	AAGTGGGGCT	9360
CTGTTGATAG	TTCTTGCTGA	GCAGAAGCCN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9420
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9480
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9540
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9600
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9660
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9720
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9780
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9840
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9900
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9960
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10020
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10080
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10140
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10200
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10260
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10320
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10380
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10440
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10500
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10560
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10620
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10680
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNGG	10740
ATCCNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10800
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10860

[illegible]

- 122 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12600
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12660
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12720
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNC	CATGGTCTAG	AACTAGTGGA	TCCCCGGGC	12780
TGCAGGAATT	CGATATCAAG	CTTATCGATA	CCGTCGACCT	CGAGGGGGGG	CCCGGTACCC	12840
AATTCGCCCT	ATAGTGAGTC	GTATTACGCG	CGCTCACTGG	CCGTCGTTTT	ACAACGTCGT	12900
GACTGGGAAA	ACCCTGGCGT	TACCCAACCT	AATCGCCTTG	CAGCACATCC	CCCTTTCGCC	12960
AGCTGGCGTA	ATAGCGAAGA	GGCCCGCACC	GATCGCCCTT	CCCAACAGTT	GCGCAGCCTG	13020
AATGGCGAAT	GGGACGCGCC	CTGTAGCGGC	GCATTAAGCG	CGGCGGGTGT	GGTGGTTACG	13080
CGCAGCGTGA	CCGCTACACT	TGCCAGCGCC	CTAGCGCCCG	CTCCTTTCGC	TTTCTTCCCT	13140
TCCTTTCTCG	CCACGTTTCGC	CGGCTTTCCC	CGTCAAGCTC	TAAATCGGGG	GCTCCCTTTA	13200
GGGTTCCGAT	TTAGTGCTTT	ACGGCACCTC	GACCCCAAAA	AACTTGATTA	GGGTGATGGT	13260
TCACGTAGTG	GGCCATCGCC	CTGATAGACG	GTTTTTCGCC	CTTTGACGTT	GGAGTCCACG	13320
TTCTTTAATA	GTGGACTCTT	GTTCCAAACT	GGAACAACAC	TCAACCTAT	CTCGGTCTAT	13380
TCTTTTGATT	TATAAGGGAT	TTTGCCGATT	TCGGCCTATT	GGTTAAAAAA	TGAGCTGATT	13440
TAACAAAAAT	TTAACGCGAA	TTTAAACAAA	ATATTAACGC	TTACAATTTA	GGTGGCACTT	13500
TTCGGGGAAA	TGTGCGCGGA	ACCCCTATTT	GTTTATTTTT	CTAAATACAT	TCAAATATGT	13560
ATCCGCTCAT	GAGACAATAA	CCCTGATAAA	TGCTTCAATA	ATATTGAAAA	AGGAAGAGTA	13620
TGAGTATTCA	ACATTTCCGT	GTCGCCCTTA	TTCCCTTTTT	TGCGGCATTT	TGCCTTCCTG	13680
TTTTTGCTCA	CCCAGAAACG	CTGGTGAAAG	TAAAAGATGC	TGAAGATCAG	TTGGGTGCAC	13740
GAGTGGGTTA	CATCGAACTG	GATCTCAACA	GCGGTAAGAT	CCTTGAGAGT	TTTCGCCCCG	13800
AAGAACGTTT	TCCAATGATG	AGCACTTTTA	AAGTTCTGCT	ATGTGGCGCG	GTATTATCCC	13860
GTATTGACGC	CGGGCAAGAG	CAACTCGGTC	GCCGCATACA	CTATTCTCAG	AATGACTTGG	13920
TTGAGTACTC	ACCAATCACA	GAAAAGCATC	TTACGGATGG	CATGACAGTA	AGAGAATTAT	13980
GCAGTGCTGC	CATAACCATG	AGTGATAACA	CTGCGGCCAA	CTTACTTCTG	ACAACGATCG	14040
GAGGACCGAA	GGAGCTAACC	GCTTTTTTGC	ACAACATGGG	GGATCATGTA	ACTCGCCTTG	14100
ATCGTTGGGA	ACCGGAGCTG	AATGAAGCCA	TACCAAACGA	CGAGCGTGAC	ACCACGATGC	14160
CTGTAGCAAT	GGCAACAACG	TTGCGCAAAC	TATTAAGTGG	CGAACTACTT	ACTCTAGCTT	14220

- 123 -

CCCGGCAACA	ATTAATAGAC	TGGATGGAGG	CGGATAAAGT	TGCAGGACCA	CTTCTGCGCT	14280
CGGCCCTTCC	GGCTGGCTGG	TTTATTGCTG	ATAAATCTGG	AGCCGGTGAG	CGTGGGTCTC	14340
GCGGTATCAT	TGCAGCACTG	GGGCCAGATG	GTAAGCCCTC	CCGTATCGTA	GTTATCTACA	14400
CGACGGGGAG	TCAGGCAACT	ATGGATGAAC	GAAATAGACA	GATCGCTGAG	ATAGGTGCCT	14460
CACTGATTAA	GCATTGGTAA	CTGTCAGACC	AAGTTTACTC	ATATATACTT	TAGATTGATT	14520
TAAAACTTCA	TTTTTAATTT	AAAAGGATCT	AGGTGAAGAT	CCTTTTTTGAT	AATCTCATGA	14580
CCAAAATCCC	TTAACGTGAG	TTTTCGTTCC	ACTGAGCGTC	AGACCCCGTA	GAAAAGATCA	14640
AAGGATCTTC	TTGAGATCCT	TTTTTTCTGC	GCGTAATCTG	CTGCTTGCAA	ACAAAAAAAC	14700
CACCGCTACC	AGCGGTGGTT	TGTTTGCCGG	ATCAAGAGCT	ACCAACTCTT	TTTCCGAAGG	14760
TAACTGGCTT	CAGCAGAGCG	CAGATACCAA	ATACTGTCCT	TCTAGTGTAG	CCGTAGTTAG	14820
GCCACCACTT	CAAGAACTCT	GTAGACCCGC	CTACATACCT	CGCTCTGCTA	ATCCTGTTAC	14880
CAGTGGCTGC	TGCCAGTGGC	GATAAGTCGT	GTCTTACCGG	GTTGGACTCA	AGACGATAGT	14940
TACCGGATAA	GGCGCAGCGG	TCGGGCTGAA	CGGGGGGTTT	GTGCACACAG	CCCAGCTTGG	15000
AGCGAACGAC	CTACACCGAA	CTGAGATACC	TACAGCGTGA	GCTATGAGAA	AGCGCCACGC	15060
TTCCCGAAGG	GAGAAAGGCG	GACAGGTATC	CGGTAAGCGG	CAGGGTCGGA	ACAGGAGAGC	15120
GCACGAGGGA	GCTTCCAGGG	GGAAACGCCT	GGTATCTTTA	TAGTCCTGTC	GGGTTTCGCC	15180
ACCTCTGACT	TGAGCGTCGA	TTTTTTGTGAT	GCTCGTCAGG	GGGGCGGAGC	CTATGGAAAA	15240
ACGCCAGCAA	CGCGGCCTTT	TTACGGTTCC	TGGCCTTTTG	CTGGCCTTTT	GCTCACATGT	15300
TCTTTCCTGC	GTTATCCCCT	GATTCTGTGG	ATAACCGTAT	TACCGCCTTT	GAGTGAGCTG	15360
ATACCGCTCG	CCGCAGCCGA	ACGACCGAGC	GCAGCGAGTC	AGTGAGCGAG	GAAGCGGAAG	15420
AGCGCCCAAT	ACGCAAACCG	CCTCTCCCCG	CGCGTTGGCC	GATTCATTAA	TGCAGCTGGC	15480
ACGACAGGTT	TCCCGACTGG	AAAGCGGGCA	GTGAGCGCAA	CGCAATTAAT	GTGAGTTAGC	15540
TCACTCATT	GGCACCCAG	GCTTTACACT	TTATGCTTCC	GGCTCGTATG	TTGTGTGGAA	15600
TTGTGAGCGG	ATAACAATTT	CACACAGGAA	ACAGCTATGA	CCATGATTAC	GCCAAGCGCG	15660
CAATTAACCC	TACTAAAGG	GAACAAAAGC	TG			15692

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15701 base pairs

(B) TYPE: nucleic acid

- 124 -

(C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = "Targetting vector"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
 (C) INDIVIDUAL ISOLATE: Swedish-FAD APP713

(ix) FEATURE:
 (A) NAME/KEY: mat_peptide
 (B) LOCATION: 4807..4983

(ix) FEATURE:
 (A) NAME/KEY: mutation
 (B) LOCATION: replace(4835, "")
 (D) OTHER INFORMATION: /standard_name= "Swedish-FAD"

(ix) FEATURE:
 (A) NAME/KEY: mutation
 (B) LOCATION: replace(4981, "")
 (D) OTHER INFORMATION: /standard_name= "APP713stop"

(ix) FEATURE:
 (A) NAME/KEY: mat_peptide
 (B) LOCATION: 8232..9032

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GAGCTCCACC GCGGTGGCGG CCGCAAGTTT AAACATGGCG CGCCTTGTCG ACGGTATCGA	60
TAAGCTTGAT CTGATGGAAT TAGAACTTGG CAAAACAATA CTGAGAATGA AGTGTATGTG	120
GAACAGAGGC TGCTGATCTC GTTCTTCAGG CTATGAAACT GACACATTTG GAAACCACAG	180
TACTTAGAAC ACAAAGTGGG AATCAAGAGA AAAACAATGA TCCCACGAGA GATCTATAGA	240
TCTATAGATC ATGAGTGGAG GAATGAGCTG GCCCTTAATT TGGTTTGTCT TGTTTAAATT	300
ATGATATCCA ACTATGAAAC ATTATCATAA AGCAATAGTA AAGAGCCTTC AGTAAAGAGC	360
AGGCATTTAT CTAATCCAC CCCACCCCA CCCCCGTAGC TCCAATCCTT CCATTCAAAA	420
TGTAGGTACT CTGTTCTCAC CTTCTTAAC AAAGTATGAC AGGAAAAAAC TTCCATTTTA	480
GTGGACATCT TTATTGTTTA ATAGATCATC AATTTCTGCA GGTCTGGCC GGGGTCCCGC	540
GGAAACTCGG CCGTGGTGAC CAATACAAA CAAAAGCGCT CCTCGTACCA GCGAAGAAGG	600
GGCAGAGATG CCGTAGTCAG GTTTAGTTCG TCCGGCGGCG CCAGAAATCC GCGCGGTGGT	660

- 125 -

TTTTGGGGGT	CGGGGGTGTT	TGGCAGCCAC	AGACGCCCCG	TGTTCTGTGC	GCGCCAGTAC	720
ATGCGGTCCA	TGCCCAGGCC	ATCCAAAAAC	CATGGGTCTG	TCTGCTCAGT	CCAGTCGTGG	780
ACCTGACCCC	ACGCAACGCC	CAAAAGAATA	ACCCCCACGA	ACCATAAACC	ATTCCCCATG	840
GGGGACCCCG	TCCCTAACCC	ACGGGGCCCC	TGGCTATGGC	GGGCTTGCCG	CCCCGACGTT	900
GGCTGCGAGC	CCTGGGCCTT	CACCCGAACT	TGGGGGTTGG	GGTGGGGAAA	AGGAAGAAAC	960
GCGGGCGTAT	TGGCCCCAAT	GGGGTCTCGG	TGGGGTATCG	ACAGAGTGCC	AGCCCTGGGA	1020
CCGAACCCCG	CGTTTATGAA	CAAACGACCC	AACACCCGTG	CGTTTTATTTC	TGTCTTTTAA	1080
TTGCCGTCAT	AGCGCGGGTT	CCTTCCGGTA	TTGTCTCCTT	CCGTGTTTCA	GTTAGCCTCC	1140
CCCATCTCCC	GGGCAAACGT	GCGCGCCAGG	TCGCAGATCG	TCGGTATGGA	GCCTGGGGTG	1200
GTGACGTGGG	TCTGGACCAT	CCCGGAGGTA	AGTTGCAGCA	GGGCGTCCCG	GCAGCCGGCG	1260
GGCGATTGGT	CGTAATCCAG	GATAAAGACG	TGCATGGGAC	GGAGGCGTTT	GGCCAAGACG	1320
TCCAAGGCCC	AGGCAAACAC	GTTATACAGG	TCGCCGTTGG	GGGCCAGCAA	CTCGGGGGCC	1380
CGAAACAGGG	TAAATAACGT	GTCCCCGATA	TGGGGTCGTG	GGCCCGCGTT	GCTCTGGGGC	1440
TCGGCACCCCT	GGGGCGGCAC	GGCCGTCCCC	GAAAGCTGTC	CCCAATCCTC	CCACCACGAC	1500
CCGCCGCCCT	GCAGATACCG	CACCGTATTG	GCAAGCAGCC	CGTAAACGCG	GCGAATCGCG	1560
GCCAGCATAG	CCAGGTCAAG	CCGCTCGCCG	GGGCGCTGGC	GTTTGGCCAG	GCGGTCTGATG	1620
TGTCTGTCTT	CCGGAAGGGC	CCCCAACACG	ATGTTTGTGC	CGGGCAAGGT	CGGCGGGATG	1680
AGGGCCACGA	ACGCCAGCAC	GGCCTGGGGG	GTCATGCTGC	CCATAAGGTA	TCGCGCGGCC	1740
GGGTAGCACA	GGAGGGCGGC	GATGGGATGG	CGGTCAAGA	TGAGGGTGAG	GGCCGGGGGC	1800
GGGGCATGTG	AACTCCCAGC	CTCCCCCCCC	ACATGAGGAG	CCAGAACGGC	GTCGGTCACG	1860
GCATAAGGCA	TGCCATTGT	TATCTGGGCG	CTTGTCATTA	CCACCGCCGC	GTCCCCGGCC	1920
GATATCTCAC	CCTGGTCGAG	GCGGTGTTGT	GTGGTGTAGA	TGTTCTCGAT	TGTCTCGGAA	1980
GCCCCCAGCA	CCTGCCAGTA	AGTCATCGGC	TCGGGTACGT	AGACGATATC	GTCGCGCGAA	2040
CCCAGGGCCA	CCAGCAGTTG	CGTGGTGGTG	GTTTTCCCCA	TCCCGTGAGG	ACCGTCTATA	2100
TAAACCCGCA	GTAGCGTGGG	CATTTTCTGC	TCCAGGCGGA	CTTCCGTGGC	TTCTTGCTGC	2160
CGGCGAGGGC	GCAACGCCGT	ACGTCGGTTG	CTATGGCCGC	GAGAACGCGC	AGCCTGGTCG	2220
AACGCAGACG	CGTGTTGATG	GCAGGGGTAC	GAAGCCATAC	GCGCTTCTAC	AAGGCGCTTG	2280
CAAAGAGGT	GCGGGAGTTT	CACGCCACCA	AGATCTGCGG	CACGCTGTTG	ACGCTGTAA	2340

- 126 -

GCGGGTCGCT	GCAGGTCGAA	AGGCCCCGAG	ATGAGGAAGA	GGAGAACAGC	GCGGCAGACG	2400
TGCGCTTTTG	AAGCGTGCAG	AATGCCGGGC	TCCGGAGGAC	CTTCGCGCCC	GCCCCGCCCC	2460
TGAGCCCGCC	CCTGAGCCCG	CCCCCGGACC	CACCCCTTCC	CAGCCTCTGA	GCCCAGAAAG	2520
CGAAGGAGCA	AAGCTGCTAT	TGGCCGCTGC	CCCAAAGGCC	TACCCGCTTC	CATTGCTCAG	2580
CGGTGCTGTC	CATCTGCACG	AGACTAGTGA	GACGTGCTAC	TTCCATTTGT	CACGTCTCTG	2640
ACGACGCGAG	CTGCGGGGCG	GGGGGGAAGT	TCCTGACTAG	GGGAGGAGTA	GAAGGTGGCG	2700
CGAAGGGGCC	ACCAAAGAAG	GGAGCCGGTT	GGCGCTACCG	GTGGATGTGG	AATGTGTGCG	2760
AGGCCAGAGG	CCACTTGTGT	AGCGCCAAGT	GCCAGCGGGG	CTGCTAAAGC	GCATGCTCCA	2820
GACTGCCTTG	GGAAAAGCGC	CTCCCCTACC	CCGGTAGAAT	TCCTGCAGCC	CGGGGGATCC	2880
ACTAGTTCTA	GAGCGGCCGC	TCTGACCATG	GNNNNNNNNN	NNNNNNNNNA	AGCTTNNNNN	2940
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3000
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3060
NNNNNNNNNN	NNNNNNNNCAT	ATGNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3120
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3180
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3240
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3300
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3360
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3420
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3480
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3540
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNGAATTCT	NNNNNNNNNN	3600
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3660
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3720
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3780
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNGAATT	CNNNNNNNNN	3840
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3900
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	3960
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4020

- 127 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4080
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNG	AATTCNNNNN	4140
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4200
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4260
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4320
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4380
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4440
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4500
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4560
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4620
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4680
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4740
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	4800
NNNGTTCTG	GGCTGACAA	CATCAAGACG	GAAGAGATCT	CTGAAGTGAA	TCTAGATGCA	4860
GAATTCCGAC	ATGACTCAGG	ATATGAAGTT	CATCATCAAA	AATTGGTGTT	CTTTGCAGAA	4920
GATGTGGGTT	CAAACAAAGG	TGCAATCATT	GGACTCATGG	TGGGCGGTGT	TGTCATAGCG	4980
TAGGGATCCA	CAGTGATCGT	CATCACCTTG	GTGATGCTGA	AGAAGAAACA	GTACACATCC	5040
ATTCATCATG	GTGTGGTGGA	GGTTGACGCC	GCTGTCACCC	CAGAGGAGCG	CCACCTGTCC	5100
AAGATGCAGC	AGAACGGCTA	CGAAAATCCA	ACCTACAAGT	TCTTTGAGCA	GATGCAGAAC	5160
TAGACCCCCG	CCACAGCAGC	CTCTGAAGTT	GGACAGCAAA	ACCATTGCTT	CACTACCCAT	5220
CGGTGTCCAT	TTATAGAATA	ATGTGGGAAG	AAACAAACCC	GTTTTATGAT	TTACTCATT	5280
TCGCCTTTTG	ACAGCTGTGC	TGTAACACAA	GTAGATGCCT	GAACCTGAAT	TAATCCACAC	5340
ATCAGTAATG	TATTCTATCT	CTCTTTACAT	TTTGGTCTCT	ATACTACATT	ATTAATGGGT	5400
TTTGTGTACT	GTAAAGAATT	TAGCTGTATC	AAACTAGTGC	ATGAATAGAT	TCTCTCCTGA	5460
TTATTTATCA	CATAGCCCCT	TAGCCAGTTG	TATATTATTC	TTGTGGTTTG	TGACCCAATT	5520
AAGTCCTACT	TTACATATGC	TTTAAGAATC	GATGGGGGAT	GCTTCATGTG	AACGTGGGAG	5580
TTCAGCTGCT	TCTCTTGCCT	AAGTATTCCT	TTCCTGATCA	CTATGCATTT	TAAAGTTAAA	5640
CATTTTAAAG	TATTTTCAGAT	GCTTTAGAGA	GATTTTTTTT	CCATGACTGC	ATTTTACTGT	5700

- 128 -

ACAGATTGCT	GCTTCTGCTA	TATTTGTGAT	ATAGGAATTA	AGAGGATACA	CACGTTTGT	5760
TCTTCGTGCC	TGTTTTATGT	GCACACATTA	GGCATTGAGA	CTTCAAGCTT	TTCTTTTTTT	5820
GTCCACGTAT	CTTTGGGTCT	TTGATAAAGA	AAAGAATCCC	TGTTCAATTGT	AAGCACTTTT	5880
ACGGGGCGGG	TGGGGAGGGG	TGCTCTGCTG	GTCTTCAATT	ACCAAGAATT	CTCCAAAACA	5940
ATTTTCTGCA	GGATGATTGT	ACAGAATCAT	TGCTTATGAC	ATGATCGCTT	TCTACACTGT	6000
ATTACATAAA	TAAATTAAAT	AAAATAACCC	CGGGCAAGAC	TTTTCTTTGA	AGGATGACTA	6060
CAGACATTAA	ATAATCGAAG	TAATTTTGGG	TGGGGAGAAG	AGGCAGATTC	AATTTTCTTT	6120
AACCAGTCTG	AAGTTTCATT	TATGATACAA	AAGAAGATGA	AAATGGAAGT	GGCAATATAA	6180
GGGGATGAGG	AAGGCATGCC	TGGACAAACC	CTTCTTTTAA	GATGTGTCTT	CAATTTGTAT	6240
AAAATGGTGT	TTTCATGTAA	ATAAATACAT	TCTTGAGGGA	GCCACATTGT	GCTGGTGTGA	6300
ATGATTCCAT	AGTAACAATC	TTGACCATTT	ACTGACGTAC	AGACCAGTGA	GAAGTCTTCG	6360
CATGTTGGGT	ACCCACACCT	GTTGTGTCTT	AATTGCAAGT	CTGAGTAGGA	AGTTGGGGCC	6420
AACATGTGTC	TCCCAGTGCT	GGGAAAATAT	TTCATAGACC	TAATTTACAG	TCTTTACTTG	6480
ATCTAAAACA	TTTTGCTGCC	ATATTTTGGC	CCTCAAGTTT	GTCCCAAATG	AGAGACAAAG	6540
GGAAAAGTTC	CAGGGAAATA	AAAATTAAGA	CAGCTGATTA	TCTGTAAAGC	ATGGTTTCTC	6600
ATCCTGAACG	CTACTAACAT	TTTGCAGGGA	ATAATTCCTT	GTTGAAGGGA	GTTGTCCTGA	6660
CCAGTGTAGG	ATATTTATTT	ATTTTATTTA	TGTTTTTTGA	GACGGAGTCT	CGTCTGTGCA	6720
CCCAGGCTGG	AGTGCAGTGG	CACAATCTCG	GCTCACTGCA	AGCTCCGCCT	CCCGGGTTCA	6780
CGCCATTCTC	CTGCCTCAGC	CTCCTGAATA	GCTGGGACTC	TAGGTGCCCC	CCACCACGCC	6840
CGGCTAATTT	TTTGTATTTT	TAGTAGAGAC	GGGGTTTCAC	CGTGTTAGCC	AGGACAGTCT	6900
TGGTCTCCTG	ACCTCGTGAT	CTGCCTGCCT	CGGCCTCCCA	AAGTGCTGAG	ATTACAGGCG	6960
TGCAAGCCGC	GCCCAGCCAG	TGCTCTCCTT	TAAAAAGTAG	CCCATTTGGCT	GGGCGCAGTG	7020
GCTCACGCCT	GTAATCCCAG	CACTTTGGGA	GGCTGAGGCG	GGTGGATCAC	GAGGTCAGGA	7080
GATCAAGAAT	ATCCTGGCCA	ATATGGTGAA	ACCCCATCTC	TACTAAAAAT	ACAAAAAATA	7140
AAAAAAAAAA	AAAAAGGCCG	GGCATGGTGG	CGGGCGCTTG	TAGTCCCAGC	TACTCAGGAG	7200
GCTGAGGCAG	GAGAATGGTG	TGCACCTGGG	AGGCGGAGGT	TGCAGTGAGC	TGAGATCGCG	7260
CCACTGCACT	CCAGCCTGGG	AGACAGAGCG	AGACTCCGTC	TCAATAAATA	AATAAATAAA	7320
TAAATAAAAG	GAGGGCCTGG	CACGAATGAC	ATGCAGGGAA	GGCAGTGAGC	AGGTGGAGGT	7380

- 129 -

CCCTGTACTC	GTTGTGGTGC	CTTATCTACC	AGGCGGTTGA	GTTGACGTCT	TTGTGGACAG	7440
AATTCGAGCT	CGGTACCCGG	GGATCCTCTA	GAGTCGACCT	TAAGGTCGAC	GGTATCGATA	7500
AGCTTGGGCT	TGAACATCGA	GCGCCAGGGC	TCCGTAAAGC	TACTAGAGCA	CAGGCGGTGC	7560
CCCAACGTCC	TGGGGCCTCT	CCACTAATAA	CGGCTACTTC	CAATTGATTG	GACGCGCCAT	7620
CTTGCCTGCC	TTATGCATAT	TCAGCGGTGA	ACTGAATATT	CATGAACGAG	GCCCCGTCCCG	7680
TCCCTCCCTC	CTTCCCCCA	CCCCCGGAAC	CCGCTCCGGA	GGACCCGAAG	GGCCCCGCCT	7740
TCATTACCGA	TGCGTAGGAC	AAACCATTTT	CCCGATGTGT	GTGGGGGGAT	ACTAATGAGA	7800
GACTTTAGCT	GAAAAATGAG	CCTGAACTCC	GAAGCTGAGT	AAAAATGGCC	TAACTTTATC	7860
CTCCGTTCTG	TAAGTCCTCG	GTTTGAGTGC	ACGGGAAACC	CGAAAGGAGG	ACGACAGGAC	7920
CAGGACATTC	TCCTCCTCCT	GTCGCGTCAG	AAAGAACACC	CAACCAGGGA	GCCGGAGCCC	7980
TAGCGTCAAC	AACTCCGCCG	CGCGCGCTCC	GTGTAGGCCG	GTGCGGGCGG	CCCCGTAGCG	8040
CAAGGGAGGG	CGGGAAAGGA	AGGGGCGGGA	CACAAGGGCG	AATCTATAAA	GGGCGTCACT	8100
CAGCCAGTTC	TCTCCTCAGA	AGCGCCGAGA	GCGCGACCGG	GACGGTTGGA	GAAGAAGGTG	8160
GCTCCCGGAA	GGGGGAGAGA	CAAACTGCCG	TAACCTCTGC	CGTTCAGGAT	CATCGAATTC	8220
CTGCAGCCAA	TATGGGATCG	GCCATTGAAC	AAGATGGATT	GCACGCAGGT	TCTCCGGCCG	8280
CTTGGGTGGA	GAGGCTATTC	GGCTATGACT	GGGCACAACA	GACAATCGGC	TGCTCTGATG	8340
CCGCCGTGTT	CCGGCTGTCA	GCGCAGGGGC	GCCCGGTTCT	TTTTGTCAAG	ACCGACCTGT	8400
CCGGTGCCCT	GAATGAACTG	CAGGACGAGG	CAGCGCGGCT	ATCGTGGCTG	GCCACGACGG	8460
GCGTTCCTTG	CGCAGCTGTG	CTCGACGTTG	TCACTGAAGC	GGGAAGGGAC	TGGCTGCTAT	8520
TGGGCGAAGT	GCCGGGGCAG	GATCTCCTGT	CATCTCACCT	TGCTCCTGCC	GAGAAAGTAT	8580
CCATCATGGC	TGATGCAATG	CGGCGGCTGC	ATACGCTTGA	TCCGGCTACC	TGCCCATTCTG	8640
ACCACCAAGC	GAAACATCGC	ATCGAGCGAG	CACGTACTION	GATGGAAGCC	GGTCTTGTCTG	8700
ATCAGGATGA	TCTGGACGAA	GAGCATCAGG	GGCTCGCGCC	AGCCGAACTG	TTCGCCAGGC	8760
TCAAGGCGCG	CATGCCCCGAC	GGCGAGGATC	TCGTCTGTGAC	CCATGGCGAT	GCCTGCTTGC	8820
CGAATATCAT	GGTGGAAAAT	GGCCGCTTTT	CTGGATTTCAT	CGACTGTGGC	CGGCTGGGTG	8880
TGGCGGACCG	CTATCAGGAC	ATAGCGTTGG	CTACCCGTGA	TATTGCTGAA	GAGCTTGGCG	8940
GCGAATGGGC	TGACCGCTTC	CTCGTGCTTT	ACGGTATCGC	CGCTCCCGAT	TCGACGCGCA	9000
TCGCCTTCTA	TCGCCTTCTT	GACGAGTTCT	TCTGAGGGGA	TCAATTCTCT	AGAGCTCGCT	9060

- 130 -

GATCAGCCTC	GACTGTGCCT	TCTAGTTGCC	AGCCATCTGT	TGTTTGCCCC	TCCCCCGTGC	9120
CTTCCTTGAC	CCTGGAAGGT	GCCACTCCCA	CTGTCCTTTC	CTAATAAAAT	GAGGAAATTG	9180
CATCGCATTG	TCTGAGTAGG	TGTCATTCTA	TTCTGGGGGG	TGGGGTGGGG	CAGGACAGCA	9240
AGGGGGAGGA	TTGGGAAGAC	AATAGCAGGC	ATGCTGGGGA	TGCGGTGGGC	TCTATGGCTT	9300
CTGAGGCGGA	AAGAACCAGC	TGGGGCTCGA	GAGATCTTCA	CAANGATAGG	AAGGAGAGGA	9360
AGTGGGGCTC	TGTTGATAGT	TCTTGCTGAG	CAGAAGCCNN	NNNNNNNNNN	NNNNNNNNNN	9420
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9480
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9540
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9600
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9660
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9720
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9780
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9840
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9900
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	9960
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10020
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10080
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10140
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10200
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10260
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10320
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10380
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10440
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10500
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10560
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10620
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10680
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10740

- 131 -

NNNNNNNGGA	TCCNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10800
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10860
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10920
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	10980
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11040
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11100
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11160
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11220
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11280
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11340
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11400
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11460
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11520
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11580
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11640
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11700
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11760
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11820
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11880
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	11940
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12000
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12060
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12120
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNGG	CGCCNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12180
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12240
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12300
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12360
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12420

- 132 -

NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12480
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12540
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12600
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12660
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	12720
NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNCC	ATGGTCTAGA	ACTAGTGGAT	12780
CCCCCGGGCT	GCAGGAATTC	GATATCAAGC	TTATCGATAC	CGTCGACCTC	GAGGGGGGGC	12840
CCGGTACCCA	ATTCGCCCTA	TAGTGAGTCG	TATTACGCGC	GCTCACTGGC	CGTCGTTTTA	12900
CAACGTCGTG	ACTGGGAAAA	CCCTGGCGTT	ACCCAACCTA	ATCGCCTTGC	AGCACATCCC	12960
CCTTTTCGCCA	GCTGGCGTAA	TAGCGAAGAG	GCCCGCACC	ATCGCCCTTC	CCAACAGTTG	13020
CGCAGCCTGA	ATGGCGAATG	GGACGCGCCC	TGTAGCGGCG	CATTAAGCGC	GGCGGGTGTG	13080
GTGGTTACGC	GCAGCGTGAC	CGCTACACTT	GCCAGCGCCC	TAGCGCCCGC	TCCTTTCGCT	13140
TTCTTCCCTT	CCTTTCTCGC	CACGTTCGCC	GGCTTTCCCC	GTCAAGCTCT	AAATCGGGGG	13200
CTCCCTTTAG	GGTTCCGATT	TAGTGCTTTA	CGGCACCTCG	ACCCCAAAAA	ACTTGATTAG	13260
GGTGATGGTT	CACGTAGTGG	GCCATCGCCC	TGATAGACGG	TTTTTCGCCC	TTTGACGTTG	13320
GAGTCCACGT	TCTTTAATAG	TGGACTCTTG	TTCCAAACTG	GAACAACACT	CAACCCTATC	13380
TCGGTCTATT	CTTTTGATTT	ATAAGGGATT	TTGCCGATTT	CGGCCTATTG	GTAAAAAAT	13440
GAGCTGATTT	AACAAAAATT	TAACGCGAAT	TTTAACAAAA	TATTAACGCT	TACAATTTAG	13500
GTGGCACTTT	TCGGGGAAAT	GTGCGCGGAA	CCCCTATTTG	TTTATTTTTC	TAAATACATT	13560
CAAATATGTA	TCCGCTCATG	AGACAATAAC	CCTGATAAAT	GCTTCAATAA	TATTGAAAAA	13620
GGAAGAGTAT	GAGTATTCAA	CATTTCCGTG	TCGCCCTTAT	TCCCTTTTTT	GCGGCATTTT	13680
GCCTTCCTGT	TTTTGCTCAC	CCAGAAACGC	TGGTGAAAGT	AAAAGATGCT	GAAGATCAGT	13740
TGGGTGCACG	AGTGGGTTAC	ATCGAACTGG	ATCTCAACAG	CGGTAAGATC	CTTGAGAGTT	13800
TTCGCCCCGA	AGAACGTTTT	CCAATGATGA	GCACTTTTAA	AGTTCTGCTA	TGTGGCGCGG	13860
TATTATCCCG	TATTGACGCC	GGGCAAGAGC	AACTCGGTCT	CCGCATACAC	TATTCTCAGA	13920
ATGACTTGGT	TGAGTACTCA	CCAGTCACAG	AAAAGCATCT	TACGGATGGC	ATGACAGTAA	13980
GAGAATTATG	CAGTGCTGCC	ATAACCATGA	GTGATAACAC	TGCGGCCAAC	TTACTTCTGA	14040
CAACGATCGG	AGGACCGAAG	GAGCTAACCG	CTTTTTTGCA	CAACATGGGG	GATCATGTAA	14100

- 133 -

CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA	14160
CCACGATGCC TGTAGCAATG GCAACAACGT TGCGCAAACCT ATTAACCTGGC GAACTACTTA	14220
CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC	14280
TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC	14340
GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG	14400
TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA	14460
TAGGTGCCTC ACTGATTAAG CATTTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT	14520
AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA	14580
ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG	14640
AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTCTGCG CGTAATCTGC TGCTTGCAAA	14700
CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT	14760
TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAA TACTGTCCTT CTAGTGTAGC	14820
CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA	14880
TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA	14940
GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCTG TGCACACAGC	15000
CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA	15060
GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA	15120
CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCTG	15180
GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC	15240
TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG	15300
CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG	15360
AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG	15420
AAGCGGAAGA GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT	15480
GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG	15540
TGAGTTAGCT CACTCATTAG GCACCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT	15600
TGTGTGGAAT TGTGAGCGGA TAACAATTTT ACACAGGAAA CAGCTATGAC CATGATTACG	15660
CCAAGCGCGC AATTAACCTT CACTAAAGGG AACAAAAGCT G	15701

(2) INFORMATION FOR SEQ ID NO:22:

- 134 -

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1297 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCATGCCTGG ACAAACCCTT CTTTAAAGAT GTGTCTTCAA TTTGTATAAA ATGGTGTTTT	60
CATGTAAATA AATACATTCT TGGAGGAGCC ACATTGTGCT GGTGTGAATG ATTCCATAGT	120
AACAATCTTG ACCATTTACT GACGTACAGA CCAGTGAGAA GTCTTCGCAT GTTGGGTACC	180
CACACCTGTT GTGTCTTAAT TGCAAGTCTG AGTAGGAAGT TGGGGCCAAC ATGTGTCTCC	240
CAGTGCTGGG AAAATATTTT ATAGACCTAA TTTACAGTCT TTAAGTGATC TAAAACATTT	300
TGCTGCCATA TTTTGCCCTT CAAGTTTGTC CCAAATGAGA GACAAAGGGA AAAGTTCCAG	360
GGAAATAAAA ATTAAGACAG CTGATTATCT GTAAAGCATG GTTTCTCATC CTGAACGCTA	420
CTAACATTTT GCAGGGAATA ATTCCTTGTT GAAGGGAGTT GTCCTGACCA GTGTAGGATA	480
TTTATTTTAT TTATTTATGT TTTTGTAGAC GGAGTCTCGC TCTGTCACCC AGGCTGGAGT	540
GCAGTGGCAC AATCTCGGCT CACTGCAAGC TCCGCCTCCC GGGTTCACGC CATTCTCCTG	600
CCTCAGCCTC CTGAATAGCT GGGACTCTAG GTGCCCGCCA CCACGCCCGG CTAATTTTTT	660
GTATTTTTTAG TAGAGACGGG GTTTCACCGT GTTAGCCAGG ACAGTCTTGG TCTCCTGACC	720
TCGTGATCTG CCTGCCTCGG CCTCCCAAAG TGCTGAGATT ACAGGCGTGC AAGCCGCGCC	780
CAGCCAGTGC TCTCCTTTTA AAAGTAGCCC ATTGGCTGGG CGCAGTGGCT CACGCCTGTA	840
ATCCCAGCAC TTTGGGAGGC TGAGGCGGGT GGATCACGAG GTCAGGAGAT CAAGAATATC	900
CTGGCCAATA TGGTGAAACC CCATCTCTAC TAAAAATACA AAAAAAAAAA AAAAAAAAAA	960
AAGGCCGGGC ATGGTGGCGG GCGCTTGTAG TCCCAGCTAC TCAGGAGGCT GAGGCAGGAG	1020
AATGGTGTGC ACCTGGGAGG CGGAGGTTGC AGTGAGCTGA GATCGCGCCA CTGCACTCCA	1080
GCCTGGGAGA CAGAGCGAGA CTCCGTCTCA ATAAATAAAT AAATAAATAA ATAAAAGGAG	1140
GGCCTGGCAC GAATGACATG CAGGGAAGGC AGTGAGCAGG TGGAGGTCCC TGTACTCGTT	1200

- 135 -

GTGGTGCCTT ATCTACCAGG CGGTTGAGTT GACGTCTTTG TGGACAGAAT TCGAGCTCGG 1260
TACCCGGGGA TCCTCTAGAG TCGACCTGCA GGCATGC 1297

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Primer"

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCATCGATGG ATCAGTTACG GAAACGATGC TCTCATGC 38

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 58 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCATCGATGG CCAAGGTGAT GACGATCACT GTGGATCCCT ACGCTATGAC AACACCGC 58

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 348 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Targetting vector"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Swedish-FAD APP

(D) OTHER INFORMATION: /standard_name= "Swedish-FAD APP"

[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26

BNSDOCID: <WO 9909150A1 | >

- 137 -

Asp Ala Glu Phe Arg His Asp Ser Gly Tyr Glu Val His His Gln Lys
 20 25 30

Leu Val Phe Phe Ala Glu Asp Val Gly Ser Asn Lys Gly Ala Ile Ile
 35 40 45

Gly Leu Met Val Gly Gly Val Val Ile Ala Thr Val Ile Val Ile Thr
 50 55 60

Leu Val Met Leu Lys Lys Lys Gln Tyr Thr Ser Ile His His Gly Val
 65 70 75 80

Val Glu Val Asp Ala Ala Val Thr Pro Glu Glu Arg His Leu Ser Lys
 85 90 95

Met Gln Gln Asn Gly Tyr Glu Asn Pro Thr Tyr Lys Phe Phe Glu Gln
 100 105 110

Met Gln Asn
 115

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 804 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Targetting vector"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..804

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ATG GGA TCG GCC ATT GAA CAA GAT GGA TTG CAC GCA GGT TCT CCG GCC	48
Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala	
1 5 10 15	
GCT TGG GTG GAG AGG CTA TTC GGC TAT GAC TGG GCA CAA CAG ACA ATC	96
Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile	
20 25 30	
GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG TCA GCG CAG GGG CGC CCG	144
Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro	
35 40 45	
GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC CTG AAT GAA CTG CAG	192

- 138 -

Val	Leu	Phe	Val	Lys	Thr	Asp	Leu	Ser	Gly	Ala	Leu	Asn	Glu	Leu	Gln	
50						55					60					
GAC	GAG	GCA	GCG	CGG	CTA	TCG	TGG	CTG	GCC	ACG	ACG	GGC	GTT	CCT	TGC	240
Asp	Glu	Ala	Ala	Arg	Leu	Ser	Trp	Leu	Ala	Thr	Thr	Gly	Val	Pro	Cys	
65					70				75						80	
GCA	GCT	GTG	CTC	GAC	GTT	GTC	ACT	GAA	GCG	GGA	AGG	GAC	TGG	CTG	CTA	288
Ala	Ala	Val	Leu	Asp	Val	Val	Thr	Glu	Ala	Gly	Arg	Asp	Trp	Leu	Leu	
				85					90					95		
TTG	GGC	GAA	GTG	CCG	GGG	CAG	GAT	CTC	CTG	TCA	TCT	CAC	CTT	GCT	CCT	336
Leu	Gly	Glu	Val	Pro	Gly	Gln	Asp	Leu	Leu	Ser	Ser	His	Leu	Ala	Pro	
			100					105					110			
GCC	GAG	AAA	GTA	TCC	ATC	ATG	GCT	GAT	GCA	ATG	CGG	CGG	CTG	CAT	ACG	384
Ala	Glu	Lys	Val	Ser	Ile	Met	Ala	Asp	Ala	Met	Arg	Arg	Leu	His	Thr	
		115					120					125				
CTT	GAT	CCG	GCT	ACC	TGC	CCA	TTC	GAC	CAC	CAA	GCG	AAA	CAT	CGC	ATC	432
Leu	Asp	Pro	Ala	Thr	Cys	Pro	Phe	Asp	His	Gln	Ala	Lys	His	Arg	Ile	
	130					135					140					
GAG	CGA	GCA	CGT	ACT	CGG	ATG	GAA	GCC	GGT	CTT	GTC	GAT	CAG	GAT	GAT	480
Glu	Arg	Ala	Arg	Thr	Arg	Met	Glu	Ala	Gly	Leu	Val	Asp	Gln	Asp	Asp	
145					150					155				160		
CTG	GAC	GAA	GAG	CAT	CAG	GGG	CTC	GCG	CCA	GCC	GAA	CTG	TTC	GCC	AGG	528
Leu	Asp	Glu	Glu	His	Gln	Gly	Leu	Ala	Pro	Ala	Glu	Leu	Phe	Ala	Arg	
				165					170					175		
CTC	AAG	GCG	CGC	ATG	CCC	GAC	GGC	GAG	GAT	CTC	GTC	GTG	ACC	CAT	GGC	576
Leu	Lys	Ala	Arg	Met	Pro	Asp	Gly	Glu	Asp	Leu	Val	Val	Thr	His	Gly	
			180					185					190			

- 139 -

GAT GCC TGC TTG CCG AAT ATC ATG GTG GAA AAT GGC CGC TTT TCT GGA	624
Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn Gly Arg Phe Ser Gly	
195 200 205	
TTC ATC GAC TGT GGC CGG CTG GGT GTG GCG GAC CGC TAT CAG GAC ATA	672
Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile	
210 215 220	
GCG TTG GCT ACC CGT GAT ATT GCT GAA GAG CTT GGC GGC GAA TGG GCT	720
Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala	
225 230 235 240	
GAC CGC TTC CTC GTG CTT TAC GGT ATC GCC GCT CCC GAT TCG CAG CGC	768
Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg	
245 250 255	
ATC GCC TTC TAT CGC CTT CTT GAC GAG TTC TTC TAG	804
Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe *	
260 265	

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 267 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala	
1 5 10 15	
Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile	
20 25 30	
Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro	
35 40 45	
Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln	
50 55 60	
Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys	
65 70 75 80	
Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu	
85 90 95	
Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro	
100 105 110	
Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr	
115 120 125	

- 140 -

Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile
 130 135 140
 Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp
 145 150 155 160
 Leu Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg
 165 170 175
 Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu Val Val Thr His Gly
 180 185 190
 Asp Ala Cys Leu Pro Asn Ile Met Val Glu Asn Gly Arg Phe Ser Gly
 195 200 205
 Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile
 210 215 220
 Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala
 225 230 235 240
 Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg
 245 250 255
 Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe
 260 265

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 348 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Targetting vector"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
 - (C) INDIVIDUAL ISOLATE: London-FAD APP
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..348
 - (D) OTHER INFORMATION: /standard_name= "London-FAD APP"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

- 141 -

TCT GGG CTG ACA AAC ATC AAG ACG GAA GAG ATC TCT GAA GTG AAG ATG	48
Ser Gly Leu Thr Asn Ile Lys Thr Glu Glu Ile Ser Glu Val Lys Met	
1 5 10 15	
GAT GCA GAA TTC CGA CAT GAC TCA GGA TAT GAA GTT CAT CAT CAA AAA	96
Asp Ala Glu Phe Arg His Asp Ser Gly Tyr Glu Val His His Gln Lys	
20 25 30	
TTG GTG TTC TTT GCA GAA GAT GTG GGT TCA AAC AAA GGT GCA ATC ATT	144
Leu Val Phe Phe Ala Glu Asp Val Gly Ser Asn Lys Gly Ala Ile Ile	
35 40 45	
GGA CTC ATG GTG GGC GGT GTT GTC ATA GCG ACA GTG ATA ATC ATC ACC	192
Gly Leu Met Val Gly Gly Val Val Ile Ala Thr Val Ile Ile Ile Thr	
50 55 60	
TTG GTG ATG CTG AAG AAG AAA CAG TAC ACA TCC ATT CAT CAT GGT GTG	240
Leu Val Met Leu Lys Lys Lys Gln Tyr Thr Ser Ile His His Gly Val	
65 70 75 80	
GTG GAG GTT GAC GCC GCT GTC ACC CCA GAG GAG CGC CAC CTG TCC AAG	288
Val Glu Val Asp Ala Ala Val Thr Pro Glu Glu Arg His Leu Ser Lys	
85 90 95	
ATG CAG CAG AAC GGC TAC GAA AAT CCA ACC TAC AAG TTC TTT GAG CAG	336
Met Gln Gln Asn Gly Tyr Glu Asn Pro Thr Tyr Lys Phe Phe Glu Gln	
100 105 110	
ATG CAG AAC TAG	348
Met Gln Asn *	

(2) INFORMATION FOR SEQ ID NO:30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 115 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Ser Gly Leu Thr Asn Ile Lys Thr Glu Glu Ile Ser Glu Val Lys Met	
1 5 10 15	
Asp Ala Glu Phe Arg His Asp Ser Gly Tyr Glu Val His His Gln Lys	
20 25 30	
Leu Val Phe Phe Ala Glu Asp Val Gly Ser Asn Lys Gly Ala Ile Ile	
35 40 45	
Gly Leu Met Val Gly Gly Val Val Ile Ala Thr Val Ile Ile Ile Thr	
50 55 60	
Leu Val Met Leu Lys Lys Lys Gln Tyr Thr Ser Ile His His Gly Val	

- 142 -

65		70		75		80									
Val	Glu	Val	Asp	Ala	Ala	Val	Thr	Pro	Glu	Glu	Arg	His	Leu	Ser	Lys
			85						90					95	
Met	Gln	Gln	Asn	Gly	Tyr	Glu	Asn	Pro	Thr	Tyr	Lys	Phe	Phe	Glu	Gln
			100					105					110		
Met	Gln	Asn													
		115													

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 348 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Targetting vector"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
 - (C) INDIVIDUAL ISOLATE: Swedish-FAD, APP
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
 - (C) INDIVIDUAL ISOLATE: London-FAD APP
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..348
 - (D) OTHER INFORMATION: /standard_name= "Swedish-FAD/London-FAD"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TCT	GGG	CTG	ACA	AAC	ATC	AAG	ACG	GAA	GAG	ATC	TCT	GAA	GTG	AAT	CTA	48
Ser	Gly	Leu	Thr	Asn	Ile	Lys	Thr	Glu	Glu	Ile	Ser	Glu	Val	Asn	Leu	
1				5				10					15			
GAT	GCA	GAA	TTC	CGA	CAT	GAC	TCA	GGA	TAT	GAA	GTT	CAT	CAT	CAA	AAA	96
Asp	Ala	Glu	Phe	Arg	His	Asp	Ser	Gly	Tyr	Glu	Val	His	His	Gln	Lys	
			20					25					30			
TTG	GTG	TTC	TTT	GCA	GAA	GAT	GTG	GGT	TCA	AAC	AAA	GGT	GCA	ATC	ATT	144
Leu	Val	Phe	Phe	Ala	Glu	Asp	Val	Gly	Ser	Asn	Lys	Gly	Ala	Ile	Ile	
			35					40					45			
GGA	CTC	ATG	GTG	GGC	GGT	GTT	GTC	ATA	GCG	ACA	GTG	ATA	ATC	ATC	ACC	192
Gly	Leu	Met	Val	Gly	Gly	Val	Val	Ile	Ala	Thr	Val	Ile	Ile	Ile	Thr	

- 143 -

50	55	60	
TTG GTG ATG CTG AAG AAG AAA CAG TAC ACA TCC ATT CAT CAT GGT GTG			240
Leu Val Met Leu Lys Lys Lys Gln Tyr Thr Ser Ile His His Gly Val			
65	70	75	80
GTG GAG GTT GAC GCC GCT GTC ACC CCA GAG GAG CGC CAC CTG TCC AAG			288
Val Glu Val Asp Ala Ala Val Thr Pro Glu Glu Arg His Leu Ser Lys			
	85	90	95
ATG CAG CAG AAC GGC TAC GAA AAT CCA ACC TAC AAG TTC TTT GAG CAG			336
Met Gln Gln Asn Gly Tyr Glu Asn Pro Thr Tyr Lys Phe Phe Glu Gln			
	100	105	110
ATG CAG AAC TAG			348
Met Gln Asn *			

(2) INFORMATION FOR SEQ ID NO:32

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 115 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Ser Gly Leu Thr Asn Ile Lys Thr Glu Glu Ile Ser Glu Val Asn Leu
 1 5 10 15
 Asp Ala Glu Phe Arg His Asp Ser Gly Tyr Glu Val His His Gln Lys
 20 25 30
 Leu Val Phe Phe Ala Glu Asp Val Gly Ser Asn Lys Gly Ala Ile Ile
 35 40 45
 Gly Leu Met Val Gly Gly Val Val Ile Ala Thr Val Ile Ile Ile Thr
 50 55 60
 Leu Val Met Leu Lys Lys Lys Gln Tyr Thr Ser Ile His His Gly Val
 65 70 75 80
 Val Glu Val Asp Ala Ala Val Thr Pro Glu Glu Arg His Leu Ser Lys
 85 90 95
 Met Gln Gln Asn Gly Tyr Glu Asn Pro Thr Tyr Lys Phe Phe Glu Gln
 100 105 110
 Met Gln Asn
 115

(2) INFORMATION FOR SEQ ID NO:33:

- 144 -

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 177 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = "Targetting vector"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 1..177
 (D) OTHER INFORMATION: /standard_name= "APP713stop"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

TCT GGG CTG ACA AAC ATC AAG ACG GAA GAG ATC TCT GAA GTG AAT CTA	48
Ser Gly Leu Thr Asn Ile Lys Thr Glu Glu Ile Ser Glu Val Asn Leu	
1 5 10 15	
GAT GCA GAA TTC CGA CAT GAC TCA GGA TAT GAA GTT CAT CAT CAA AAA	96
Asp Ala Glu Phe Arg His Asp Ser Gly Tyr Glu Val His His Gln Lys	
20 25 30	
TTG GTG TTC TTT GCA GAA GAT GTG GGT TCA AAC AAA GGT GCA ATC ATT	144
Leu Val Phe Phe Ala Glu Asp Val Gly Ser Asn Lys Gly Ala Ile Ile	
35 40 45	
GGA CTC ATG GTG GGC GGT GTT GTC ATA GCG TAG	177
Gly Leu Met Val Gly Gly Val Val Ile Ala *	
50 55	

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 58 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Ser Gly Leu Thr Asn Ile Lys Thr Glu Glu Ile Ser Glu Val Asn Leu	
1 5 10 15	
Asp Ala Glu Phe Arg His Asp Ser Gly Tyr Glu Val His His Gln Lys	
20 25 30	
Leu Val Phe Phe Ala Glu Asp Val Gly Ser Asn Lys Gly Ala Ile Ile	
35 40 45	

- 145 -

Gly Leu Met Val Gly Gly Val Val Ile Ala
 50 55

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 99 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mus musculus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

TCT GGG CTG ACA AAC ATC AAG ACG GAA GAG ATC TCT GAA GTG AAG ATG	48
Ser Gly Leu Thr Asn Ile Lys Thr Glu Glu Ile Ser Glu Val Lys Met	
1 5 10 15	
GAT GCA GAA TTC CGA CAT GAC TCA GGA TAT GAA GTT CAT CAT CAA AAA	96
Asp Ala Glu Phe Arg His Asp Ser Gly Tyr Glu Val His His Gln Lys	
20 25 30	
CTG	
Leu	99

(2) INFORMATION FOR SEQ ID NO:36

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Ser Gly Leu Thr Asn Ile Lys Thr Glu Glu Ile Ser Glu Val Lys Met	
1 5 10 15	
Asp Ala Glu Phe Arg His Asp Ser Gly Tyr Glu Val His His Gln Lys	
20 25 30	
Leu	

What is claimed:

1. A recombinant nucleic acid molecule comprising:
a nucleotide sequence which is effective to achieve homologous recombination at a predefined position of a rodent APP gene, which is operably linked to the 5' terminus of a nucleotide coding sequence which, when inserted into said APP gene, codes for at least one amino acid whose identity and/or position is not naturally-occurring in said APP gene,
and a nucleotide sequence which is effective to achieve homologous recombination at a predefined position of a rodent APP gene, which is operably linked to the 3' terminus of said nucleotide coding sequence.
2. A recombinant nucleic acid molecule of claim 1, wherein said nucleotide coding sequence comprises at least one amino acid coded for by a human APP gene.
3. A recombinant nucleic acid molecule of claim 1, wherein said nucleotide coding sequence codes without interruption for an amino acid sequence, which amino acid sequence is coded for by two or more exons in a naturally-occurring genomic sequence.
4. A recombinant nucleic acid molecule of claim 1, wherein said nucleotide coding sequence codes without interruption for an amino acid

- 147 -

sequence of a human APP polypeptide, which amino acid sequence is coded for by two or more exons in a human APP genomic sequence.

5. A recombinant nucleic acid molecule of claim 1, wherein said nucleotide coding sequence hybridizes under stringent conditions to a nucleotide sequence coding without interruption for at least one amino acid sequence coded for by a human APP gene.
6. A recombinant nucleic acid molecule of claim 1, wherein said nucleotide coding sequence, which when inserted into a rodent APP gene, comprises an asparagine at amino acid position 670, a leucine at amino acid position 671, an arginine at amino acid position 676, a threonine at amino acid position 681, a histidine at amino acid 684, an isoleucine at amino acid 717, or a combination thereof.
7. A recombinant nucleic acid molecule of claim 1, wherein said nucleotide coding sequence, which when inserted into a rodent APP gene, does not comprise an asparagine at amino acid position 670, a leucine at amino acid position 671, an arginine at amino acid position 676, a threonine at amino acid position 681, and a histidine at amino acid 684.
8. A recombinant nucleic acid molecule of claim 1, further comprising a selectable marker gene.
9. A recombinant nucleic acid molecule of claim 8, wherein the selectable marker gene codes for neomycin resistance.
10. A recombinant nucleic acid molecule of claim 1, further comprising a vector.

- 148 -

11. A recombinant nucleic acid molecule of claim 10, further comprising a selectable marker gene, wherein said vector contains only one selectable marker gene which can be selected for in a mammalian host.
12. A recombinant nucleic acid molecule of claim 1, further comprising a selectable marker gene and a vector, and
wherein said nucleotide coding sequence codes without interruption for an amino acid sequence, which amino acid sequence is coded for by two or more exons in a human APP genomic sequence.
13. A recombinant nucleic acid molecule of claim 1, wherein the nucleotide coding sequence is at least part of a cDNA.
14. A recombinant nucleic acid molecule of claim 1, wherein said nucleotide sequence operably linked to the 5' terminus of said nucleotide coding sequence comprises all or part of intron 15 of a mouse APP gene.
15. A recombinant nucleic acid molecule of claim 1, wherein said nucleotide sequence operably linked to the 3' terminus of said nucleotide coding sequence comprises all or part of intron 16 of a mouse APP gene.
16. A recombinant nucleic acid molecule comprising:
a nucleotide sequence of a rodent APP gene which is operably linked to the 5' terminus of a nucleotide coding sequence which, when inserted into a mouse APP gene, codes for at least one amino acid whose identity and/or position is not naturally-occurring in said APP gene, and
a nucleotide sequence of a rodent APP gene which is operably linked to the 3' terminus of said nucleotide coding sequence coding,

- 149 -

whereby the nucleic acid molecule is effective to achieve homologous recombination in a rodent chromosome.

17. A recombinant nucleic acid coding for a humanized rodent APP polypeptide comprising at least one amino acid coded for by a human APP gene.

18. A recombinant nucleic acid of claim 17 which codes without interruption for said humanized rodent APP polypeptide, and which comprises two or more amino acids coded for by a human APP gene, wherein said amino acids of the human APP gene are coded for by two or more exons in a human APP genomic sequence.

19. A recombinant nucleic acid of claim 18, wherein the rodent is a mouse, comprising an asparagine at amino acid position 670, a leucine at amino acid position 671, an arginine at amino acid position 676, a threonine at amino acid position 681, a histidine at amino acid 684, an isoleucine at amino acid 717, or a combination thereof.

20. A humanized mouse APP polypeptide coded for by a recombinant nucleic acid of claim 19.

21. A transformed non-human mammal cell comprising a recombinant nucleic acid of claim 1.

22. A transformed mouse cell comprising a recombinant nucleic acid of claim 1.

- 150 -

23. A transformed mouse cell of claim 22, wherein the mouse cell is an embryonic stem cell.

24. A transformed mouse cell comprising a recombinant nucleic acid of claim 1, wherein the nucleic acid is integrated into the chromosome of the mouse cell at the mouse APP gene locus.

25. A transformed mouse cell comprising a recombinant nucleic acid of claim 4.

26. A transformed mouse cell comprising a recombinant nucleic acid of claim 5.

27. A transformed mouse cell comprising a recombinant nucleic acid of claim 6.

28. A transformed mouse cell comprising a recombinant nucleic acid of claim 12.

29. A transgenic rodent comprising cells which contain a recombinant APP gene integrated into a chromosome of said cell at the APP gene locus, said APP gene comprising a nucleotide coding sequence which codes for a recombinant APP polypeptide comprising at least one amino acid whose identity and/or position is not naturally-occurring in said rodent APP gene.

30. A transgenic rodent comprising cells which contain a recombinant APP gene comprising a nucleotide coding sequence for a recombinant APP polypeptide, said APP polypeptide comprising two or more amino acids coded for by a human APP gene, wherein said amino acids of the human

- 151 -

APP gene are coded for by two or more exons in a human APP genomic sequence.

31. A transgenic rodent of claim 30 which is a mouse, wherein said APP polypeptide comprises an asparagine at amino acid position 670, a leucine at amino acid position 671, an arginine at amino acid position 676, a threonine at amino acid position 681, a histidine at amino acid 684, an isoleucine at amino acid 717, or a combination thereof.
32. A transgenic rodent of claim 30, wherein said recombinant APP gene further comprises a selectable marker gene.
33. A transgenic mouse of claim 32, wherein said selectable marker gene codes for neomycin resistance.
34. A transgenic mouse of claim 30, wherein said recombinant APP gene is integrated into the rodent chromosome at the APP gene locus.
35. A transgenic mouse of claim 30, wherein the recombinant APP gene is expressed in the brain of said mouse in an amount effective to produce neuronal cell degeneration and/or apoptosis.
36. A transgenic mouse of claim 30, wherein the recombinant APP gene is expressed in the brain of said mouse in an amount effective to cause a behavioral or cognitive dysfunction, wherein the dysfunction is conferred by said polypeptide coded for by said recombinant APP gene.

- 152 -

37. A transgenic mouse of claim 29, wherein the recombinant APP gene is expressed in the brain of said rodent in an amount effective to produce neuronal cell degeneration and/or apoptosis.
38. A transgenic rodent of claim 29, wherein the recombinant APP gene is expressed in the brain of said rodent in an amount effective to cause a behavioral or cognitive dysfunction, wherein the dysfunction is conferred by said polypeptide coded for by said recombinant APP gene.
39. A method for producing a transgenic rodent comprising a recombinant nucleic acid molecule, comprising:
- (a) introducing a nucleic acid molecule of claim 1 into mouse ES cells;
 - (b) culturing said ES cells under conditions effective for homologous recombination between said nucleic acid and an APP gene of said ES cells;
 - (c) selecting cells having a nucleic acid of claim 1 integrated by homologous recombination into said APP gene of said ES cells;
 - (d) introducing said transformed ES cells into a blastocyst;
 - (e) implanting said blastocyst into a pseudopregnant rodent; and
 - (f) allowing said embryo to develop to term.
40. A method of claim 39, wherein only one APP gene targeting event of steps (a)-(c) is accomplished.
41. A transgenic rodent produced according to claim 39.
42. A method of claim 39, further comprising identifying at least one transgenic offspring containing said recombinant DNA.

- 153 -

43. A method of claim 39, further comprising breeding said offspring to form a transgenic line of rodents.
44. A method of claim 39, wherein said nucleotide coding sequence codes without interruption for an amino acid sequence, and wherein the amino acid sequence is coded for by two or more exons in a human APP genomic sequence.
45. A method for producing a transgenic rodent having a phenotype mediated by expression of a recombinant APP gene in the brain of said rodent, wherein the APP gene is expressed in the brain of said rodent in an amount effective to produce neuronal cell degeneration and/or apoptosis and/or in an amount effective to cause a behavioral or cognitive dysfunction, said method comprising:
- (a) introducing a nucleic acid molecule of claim 1 into rodent ES cells;
 - (b) culturing said ES cells under conditions effective for homologous recombination between said nucleic acid and an APP gene of said ES cells;
 - (c) selecting cells having a nucleic acid of claim 1 integrated into
 - (d) introducing said transformed ES cells into a blastocyst;
 - (e) implanting said blastocyst into a pseudopregnant rodent; and
 - (f) allowing said embryo to develop to term.
46. A method of claim 45, wherein said nucleotide coding sequence codes without interruption for an amino acid sequence, wherein the amino acid sequence is coded for by two or more exons in a human APP genomic sequence.

- 154 -

47. A method of screening a compound for an effect on a phenotype mediated by expression of a recombinant APP gene in the brain of a rodent, comprising:

administering said compound to a rodent of claim 30 expressing said recombinant APP gene, and

observing whether an effect on said phenotype results.

48. A recombinant nucleic acid molecule comprising:

a nucleotide sequence which is effective to achieve homologous recombination at a predefined position of a target gene, operably linked to the 5' terminus of a nucleotide coding sequence which codes without interruption for an amino acid sequence, which when inserted into said target gene, codes for at least two amino acids whose identity and/or position is not naturally-occurring in said target gene, wherein said amino acids are coded for by two or more exons in a genomic sequence, and

a nucleotide sequence which is effective to achieve homologous recombination at a predefined position of a non-human gene, operably linked to the 3' terminus of said nucleotide coding sequence.

49. A recombinant nucleic acid of claim 48, wherein the non-human mammal is a mouse.

50. A method for producing a transgenic rodent comprising a recombinant nucleic acid molecule, comprising:

(a) introducing a nucleic acid molecule of claim 48 into rodent ES cells;

(b) culturing said ES cells under conditions effective for homologous recombination between said nucleic acid and a target gene of said ES cells;

- 155 -

- (c) selecting cells having a nucleic acid of claim 48 integrated by homologous recombination into said target gene of said ES cells;
- (d) introducing said transformed ES cells into a blastocyst;
- (e) implanting said blastocyst into a pseudopregnant mouse; and
- (f) allowing said embryo to develop to term.

51. A method of claim 50, wherein steps (a)-(c) are accomplished only once.

52. A recombinant nucleic acid molecule comprising:

a nucleotide sequence which is effective to achieve homologous recombination at a predefined position of a target gene, operably linked to the 5' terminus of a nucleotide coding sequence which codes without interruption for an amino acid sequence, which when inserted into said target gene, codes for at least one amino acid whose identity and/or position is not naturally-occurring in said target gene, and

a nucleotide sequence which is effective to achieve homologous recombination at a predefined position of a target gene, operably linked to the 3' terminus of said nucleotide coding sequence.

53. A nucleic acid molecule of claim 52, wherein the target gene is a mammalian gene.

54. A method for producing a transgenic rodent comprising a recombinant nucleic acid molecule, comprising:

- (a) introducing a nucleic acid molecule of claim 52 into rodent ES cells;

- 156 -

(b) culturing said ES cells under conditions effective for homologous recombination between said nucleic acid and a target gene of said ES cells;

(c) selecting cells having a nucleic acid of claim 52 integrated by homologous recombination into said target gene of said ES cells;

(d) introducing said transformed ES cells into a blastocyst;

(e) implanting said blastocyst into a pseudopregnant mouse; and

(f) allowing said embryo to develop to term,

wherein only steps (a)-(c) are accomplished only once.

55. A recombinant nucleic acid molecule of claim 1, which is pMTI-2455, pMTI-2454, pMTI-2453, or pMTI-2398.

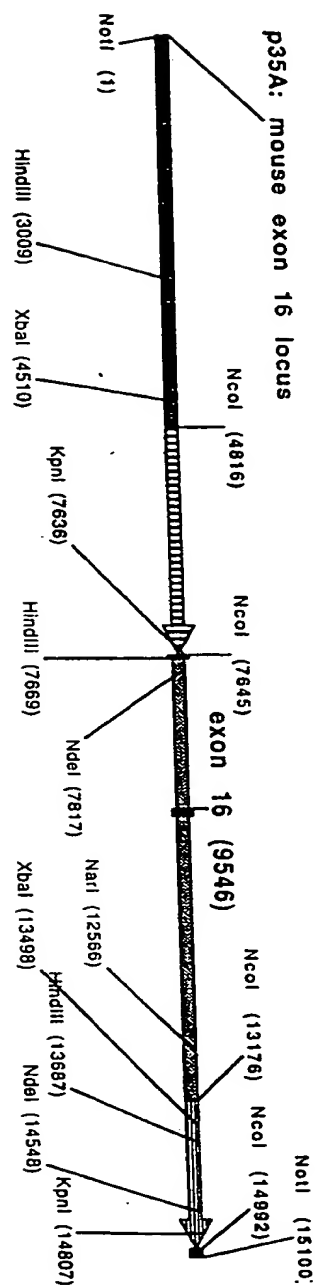


Figure 1

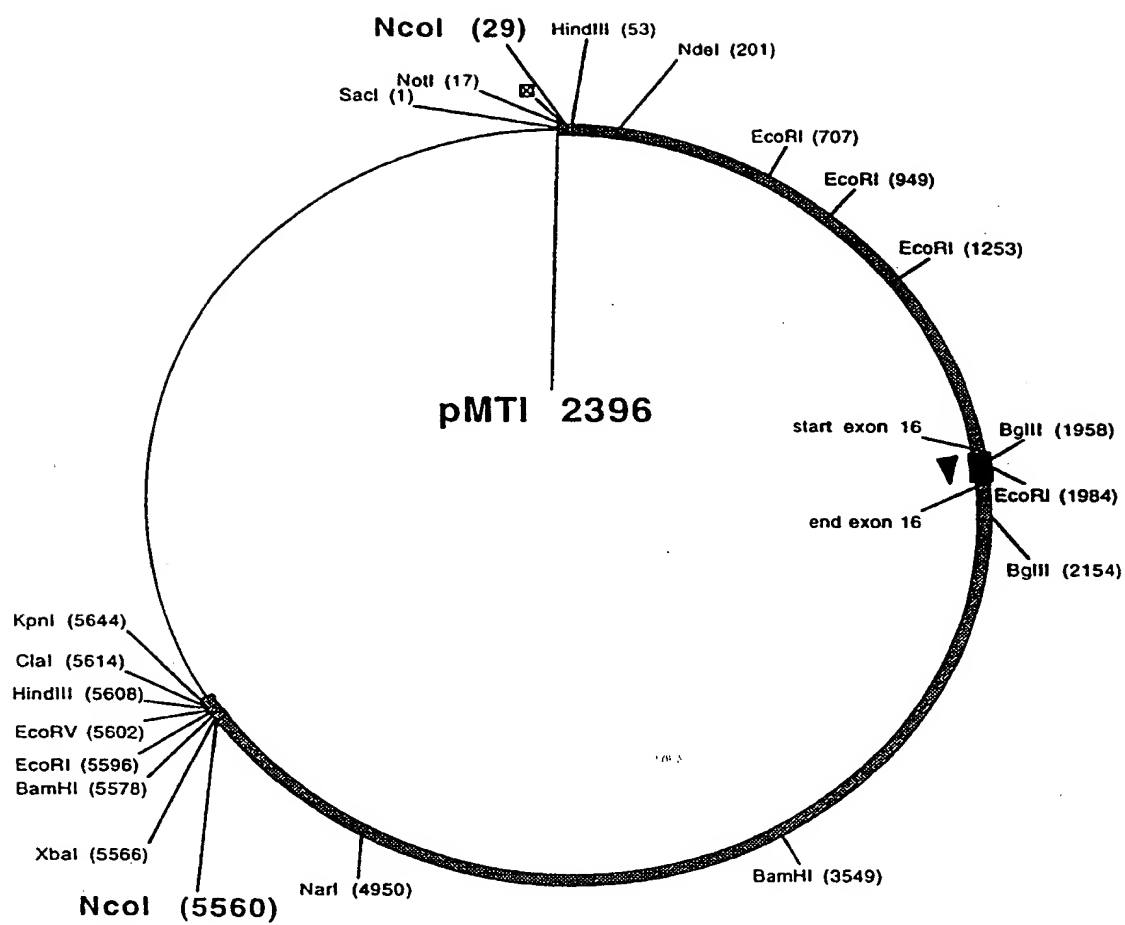


Figure 2

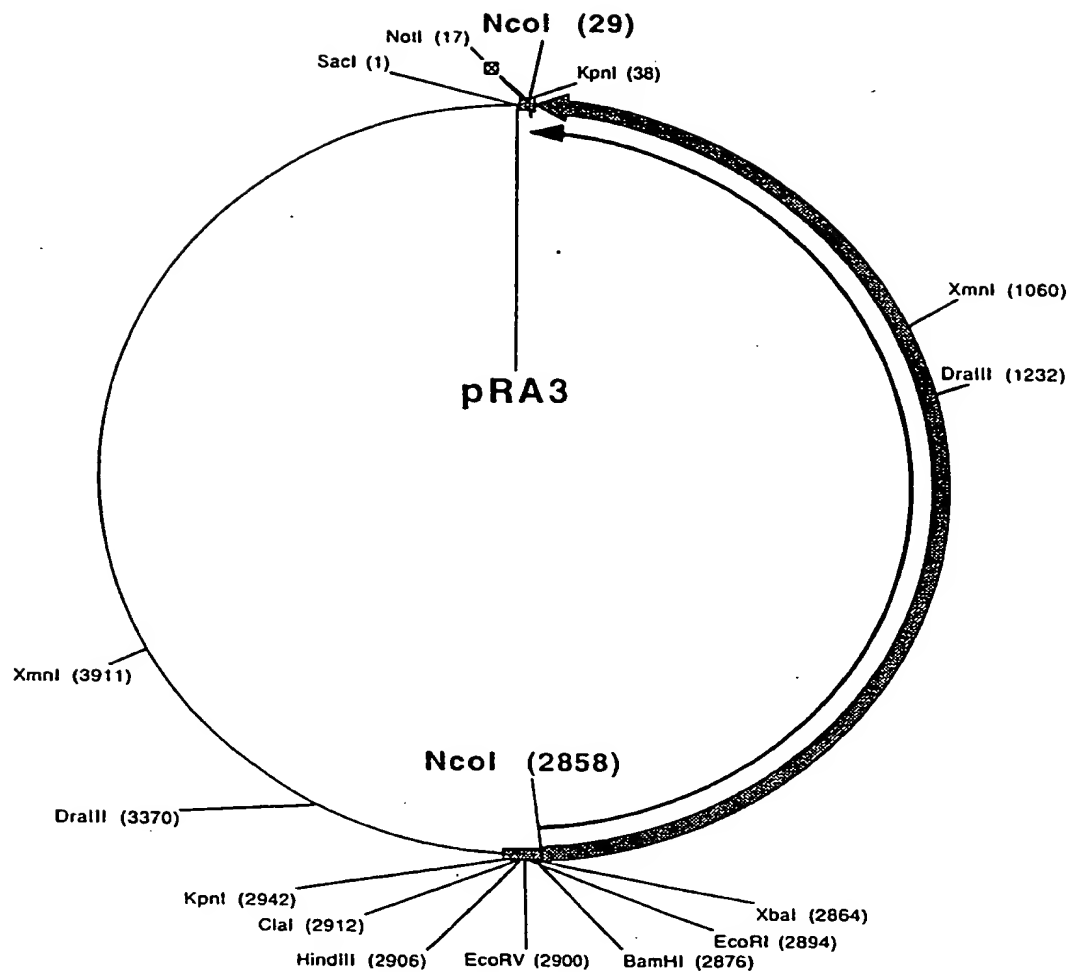


Figure 3

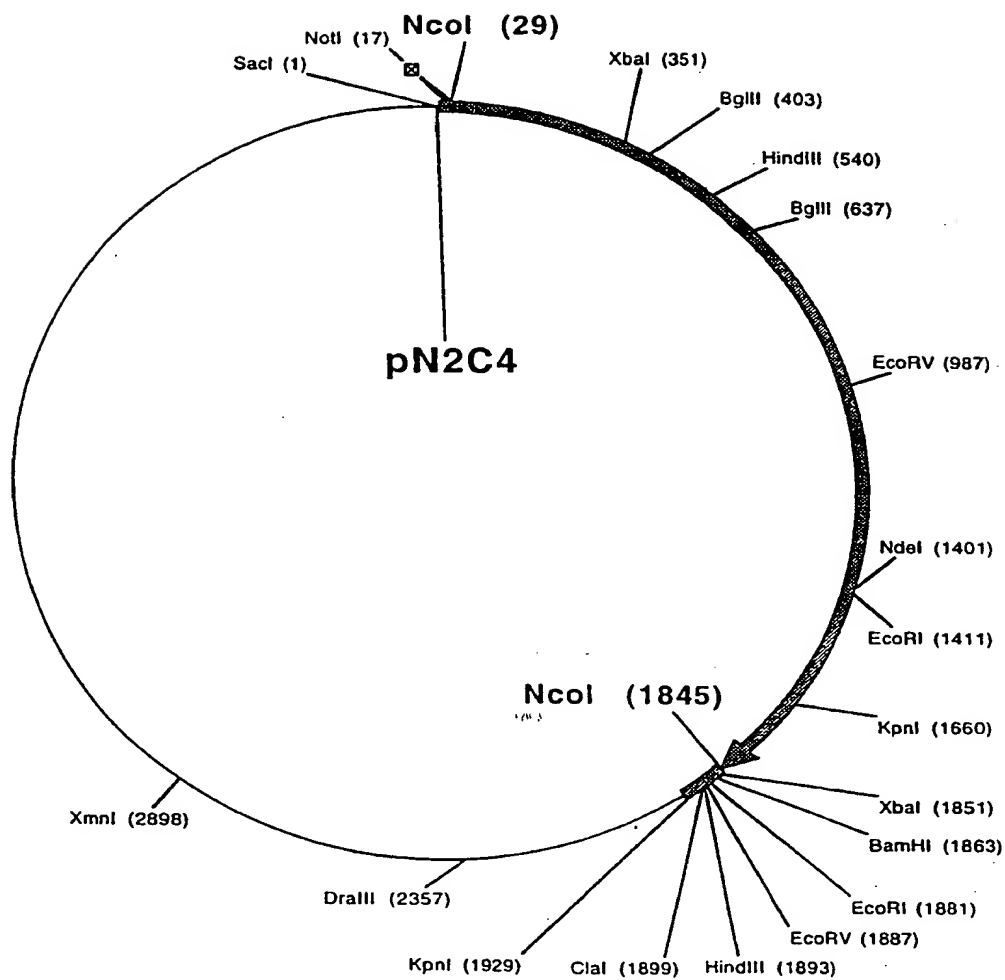


Figure 4

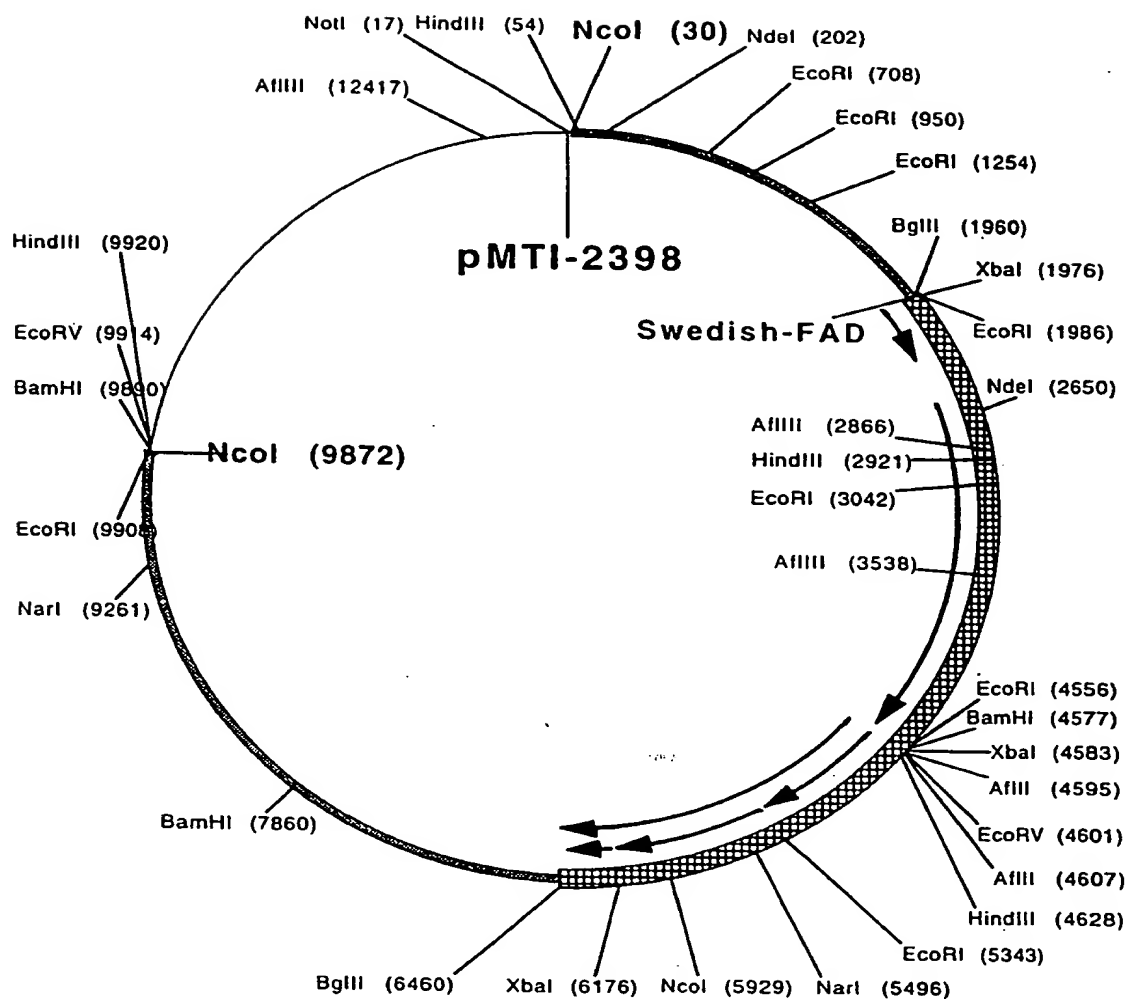


Figure 5

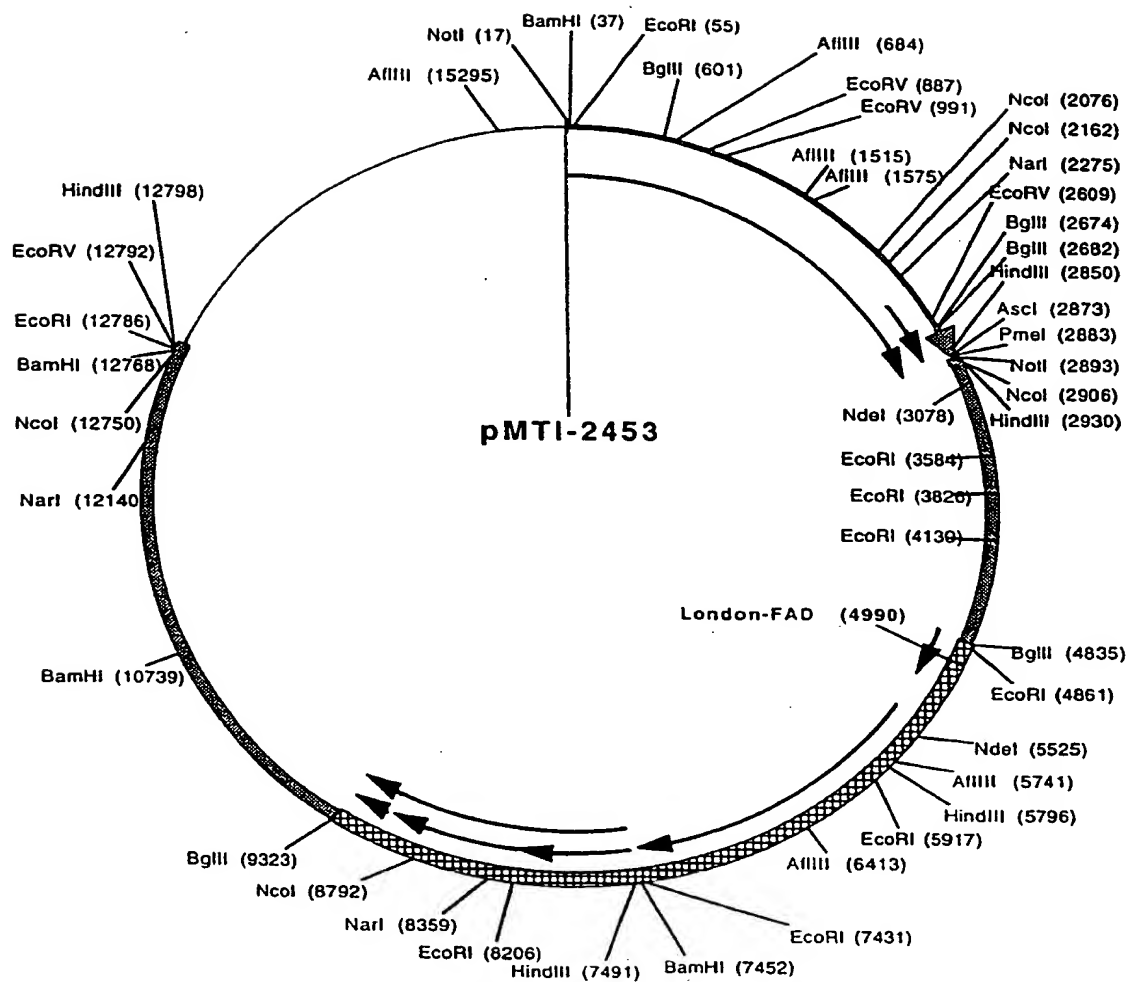


Figure 6

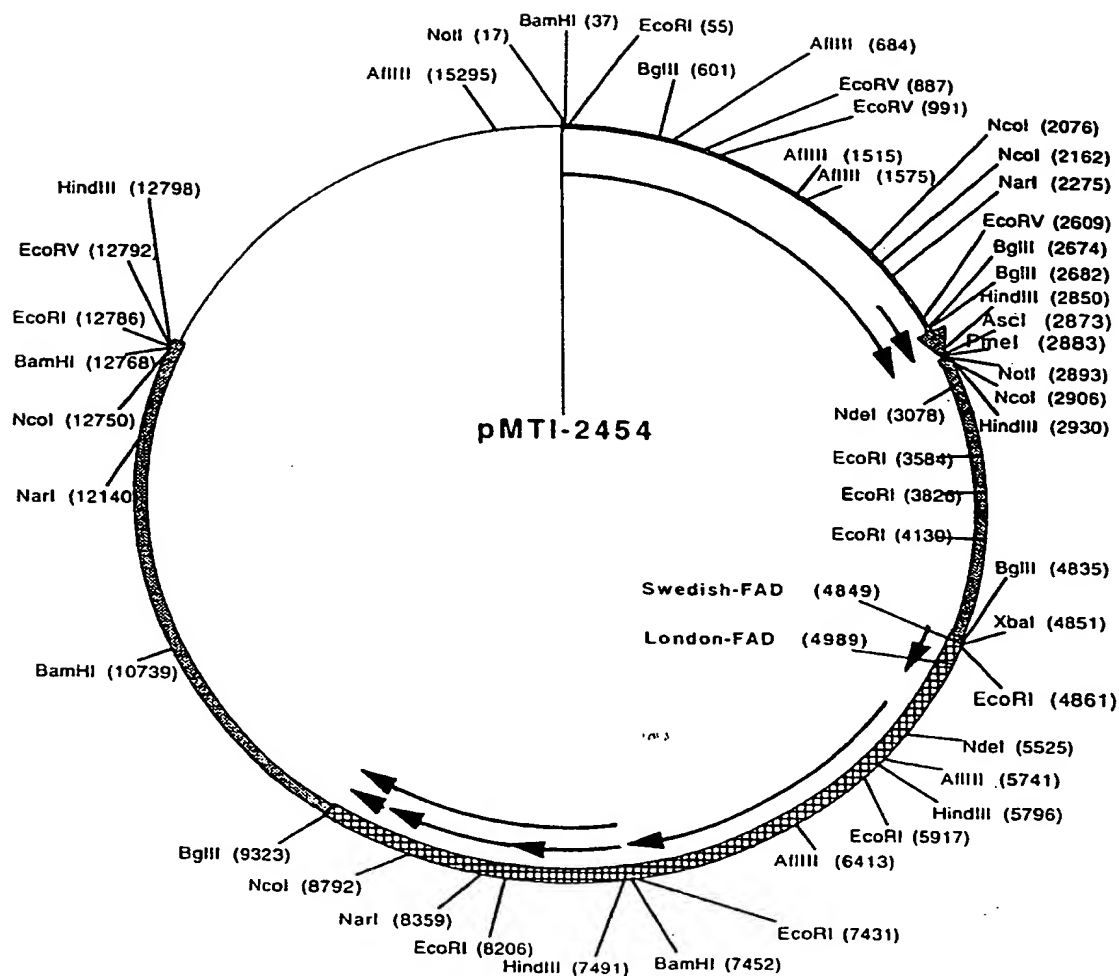


Figure 7

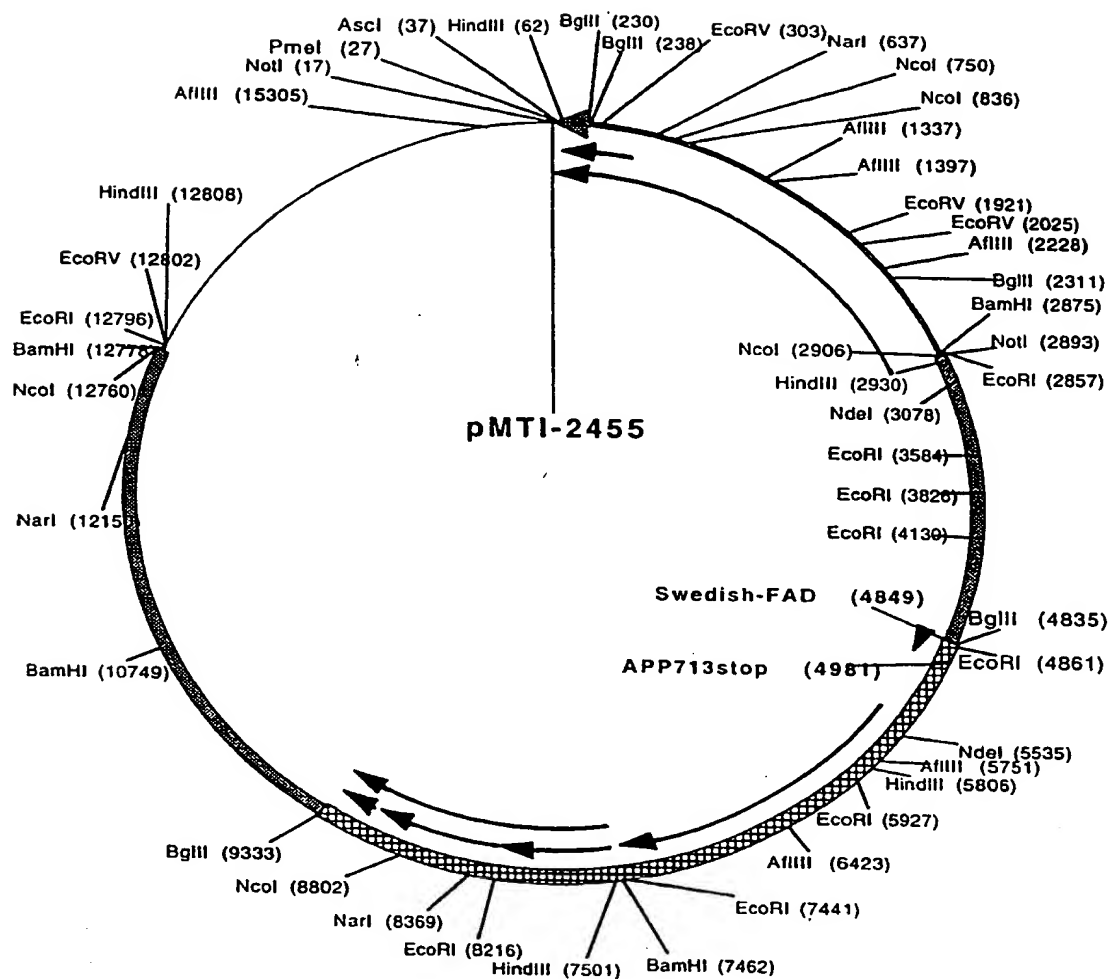


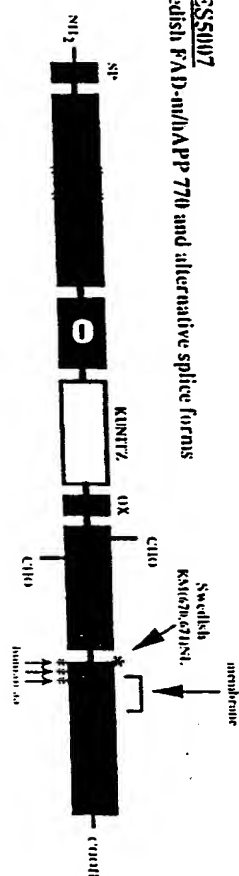
Figure 8

Figure 9

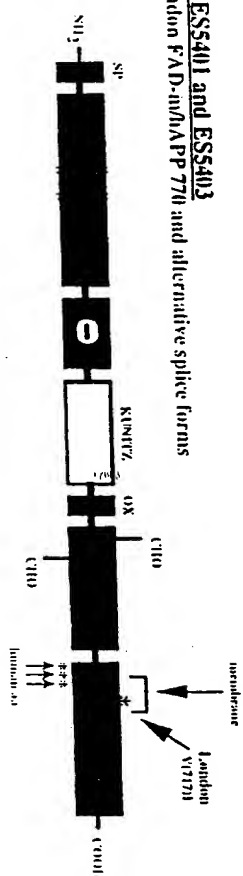
Oligonucleotide	DNA Sequence
Sall-AflII-EcoRV-NcoI-MluI adaptor	5'TCGACGACTTAAGTTGATATCCACCATGGTGACGCGTT3'
XhoI-BglII-StuI adaptor	5'TCGAGTGAGATCTTAAGGCCTGG3'
Sall-Ascl-PmeI-NotI-Ascl-PmeI-Sall adaptor	5'TCGACAAGGCGCGCCGTTTAAACAAGCGGCCGCTTGGCGCGCC TTTTGTTTAAACTTG3'
Oligonucleotide 6	5'CCTCGGCCTTTGGTGTGTGTTTTATGACATGACCCCTTGA3'
Oligonucleotide 7	5'CACCCTGTTGTCAATGCCTCTGGGTTTCCGCCAGTTTCG3'
RA49	5'CGATGGGTAGTGAAGCA3'
KC56	5'GTGAAGATGGATGCAGAATTC3'
KC65	5'GTTCTGGGCTGACAAACATC3'
KC66	5'GATGGCGGACTTCAAATCCTG3'
KC95/96	5'CTAGACACTC3'
KC125	5'ACTTTGTGTTTGACGC3'
KC131	5'GATGATGAACCTCATATCCTG3'
KC132	5'CAGTTTTTGATGGCGG3'
KC137	5'GTTTGAGACCTTCAACACCC3'
KC138	5'GAAGGAAGGCTGGAAAAGAGCC3'

Figure 9

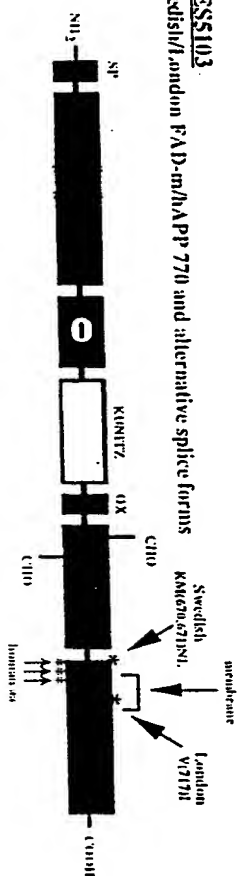
Transgenic Line ES5007
Gene Products: Swedish FAD-in/hAPP 770 and alternative splice forms



Transgenic Lines ES5401 and ES5403
Gene Products: London FAD-mAPP 77k and alternative splice forms

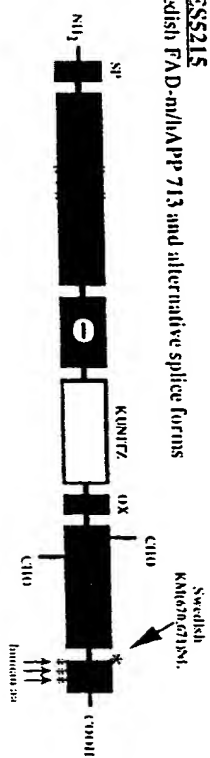


Transgenic Line ES5103
Gene Products: Swedish/ London FAD-m/hAPP 770 and alternative splice forms



Transgenic Line ES5215

Gene Products: Swedish FAD-m/IA PP 713 and alternative splice forms



14W-212-101695a

Figure 10

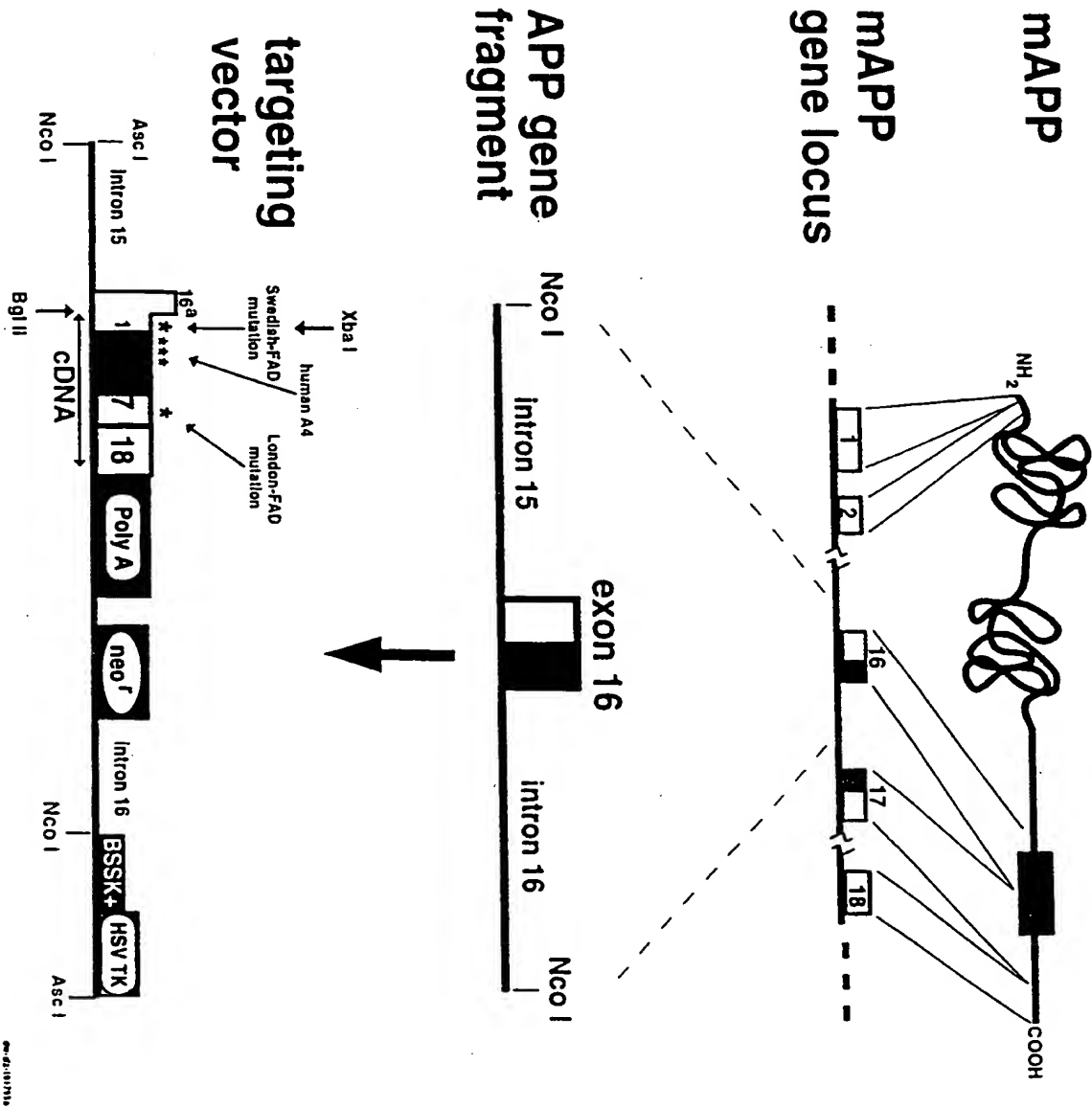


Figure 11

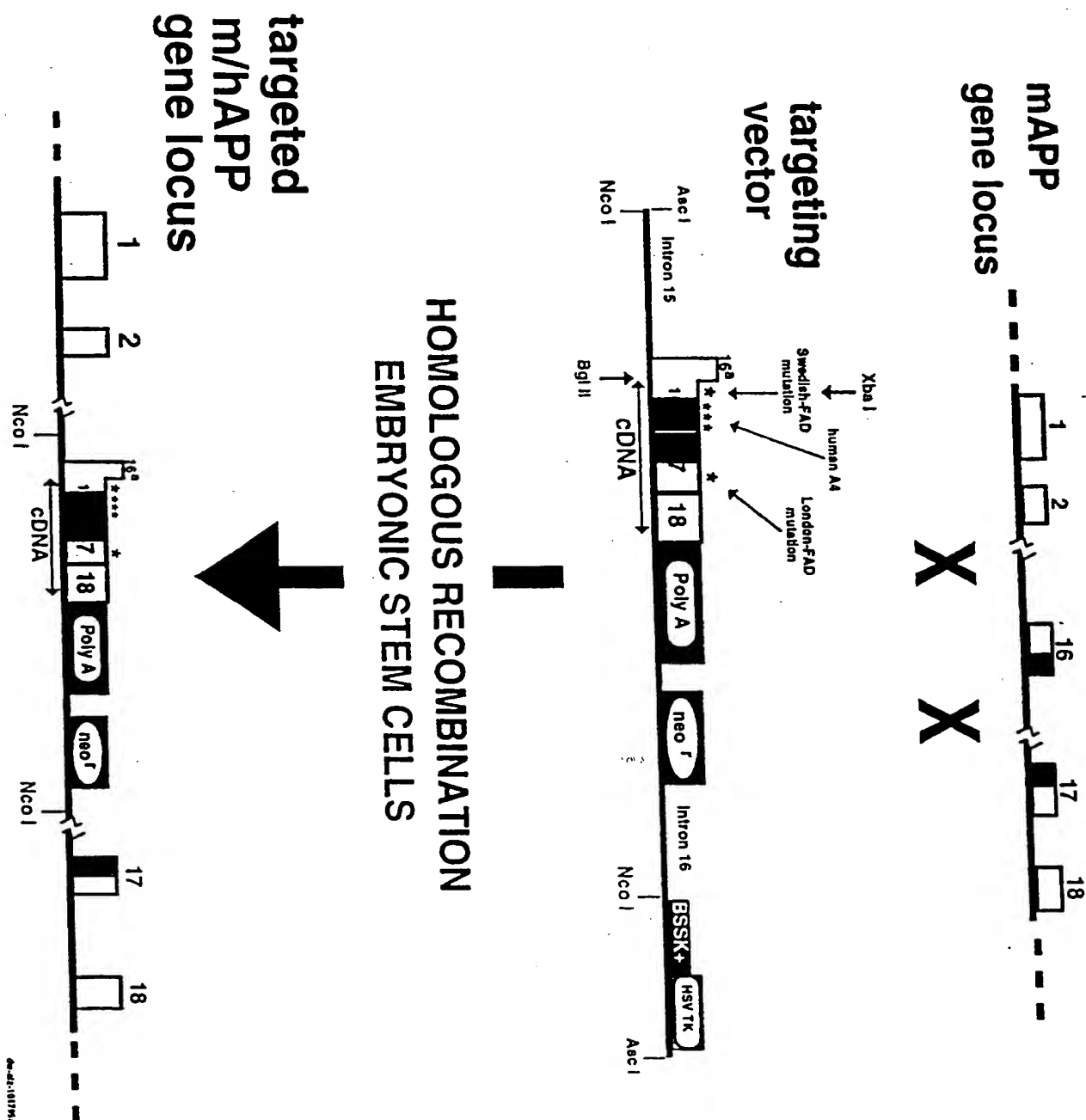


Figure 12

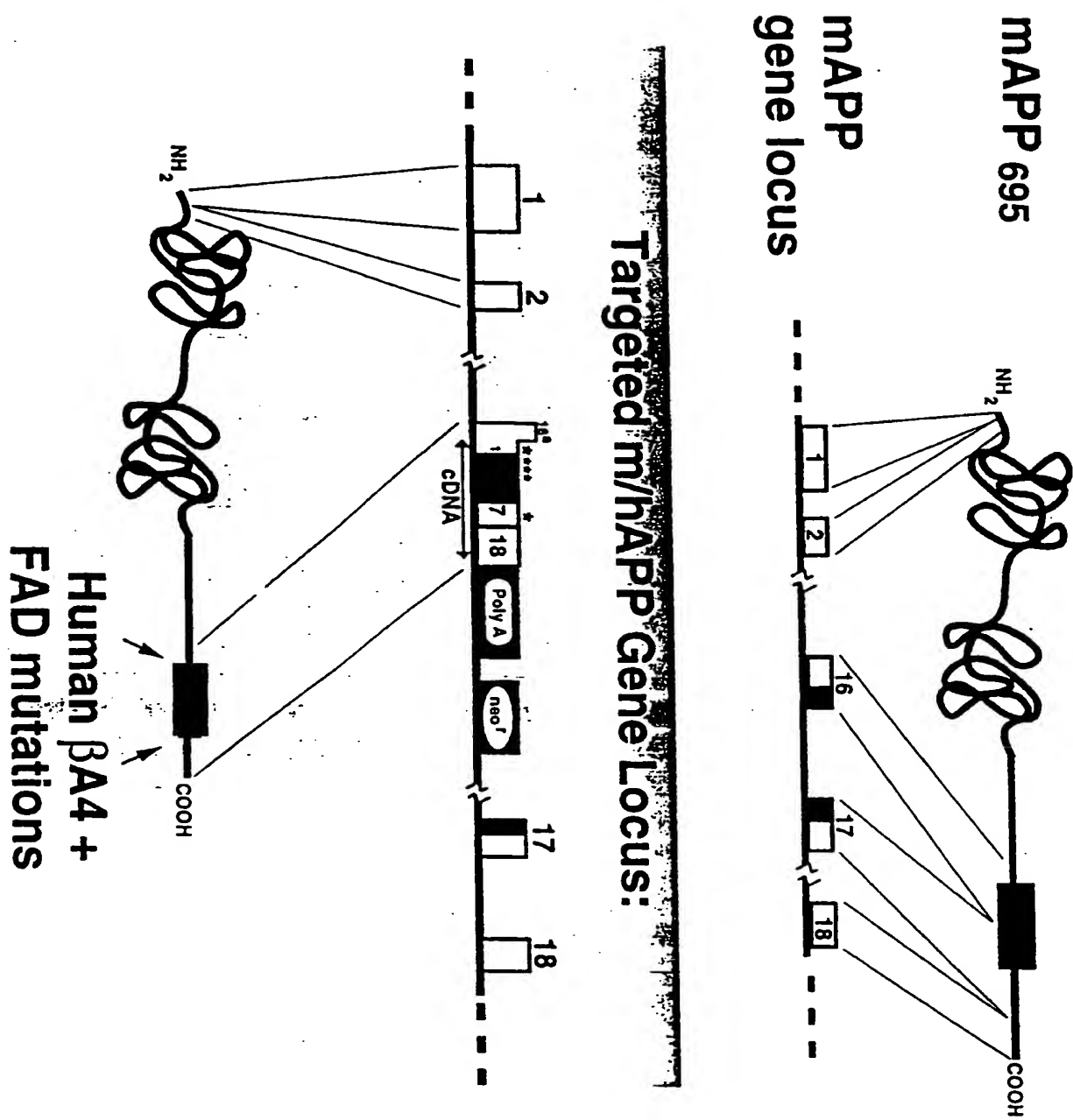


Figure 13

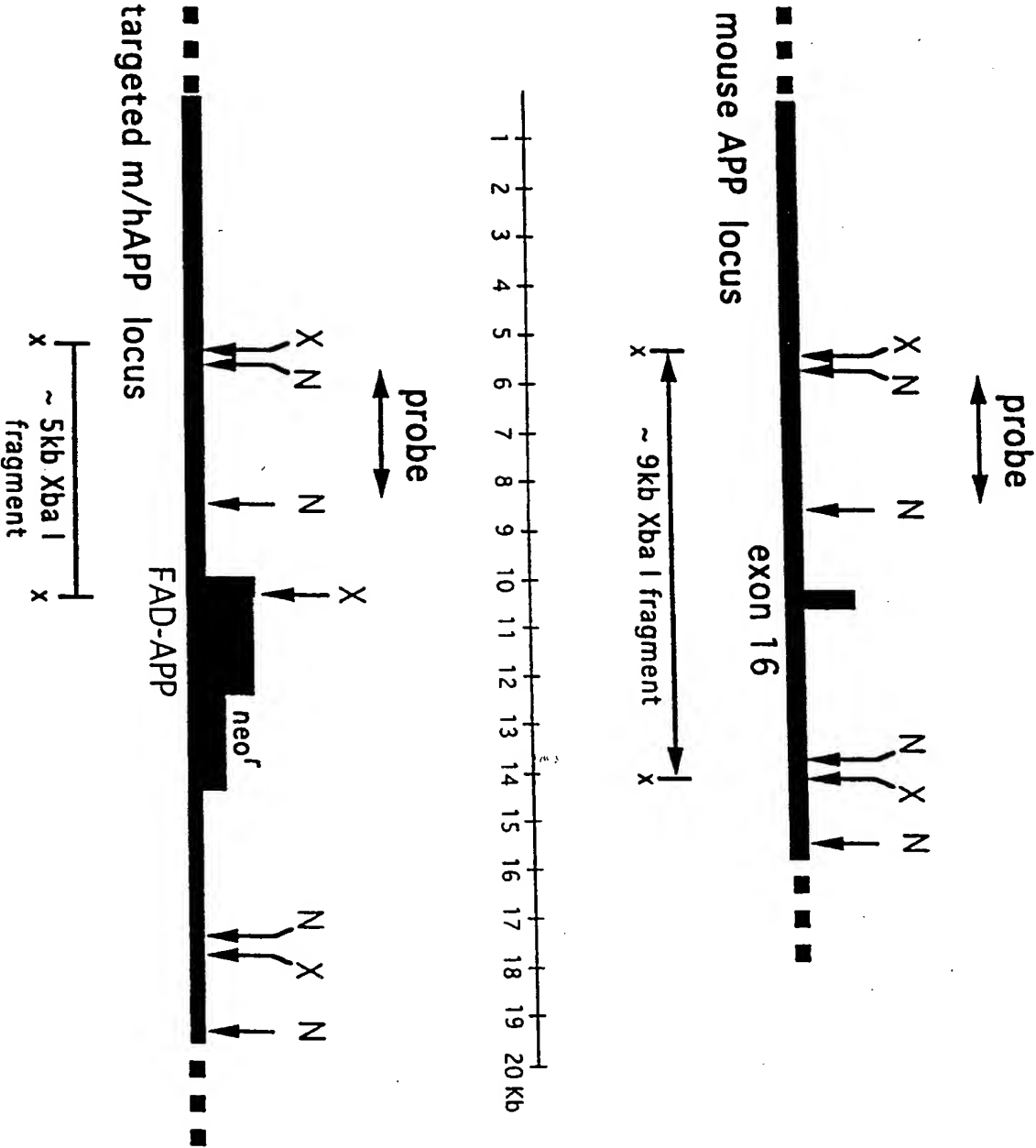
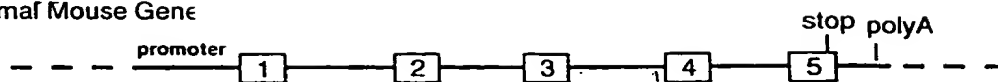
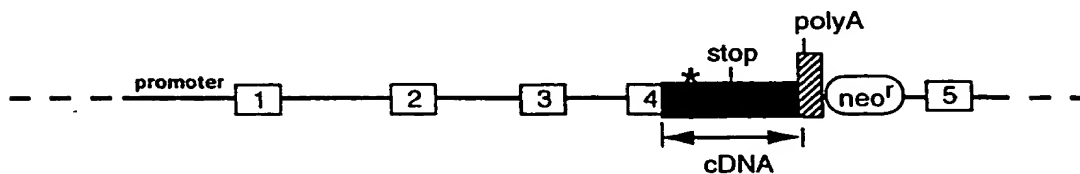


Figure 14

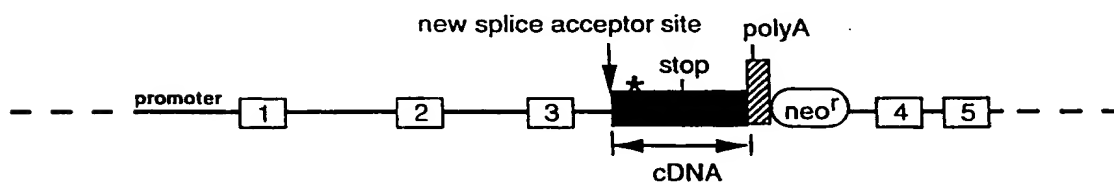
A. Normal Mouse Gene



B.



C.



D.

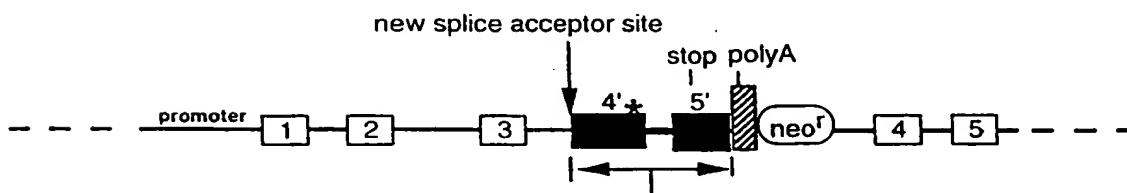


Figure 15

dw-atz-53096a

man App770.Pep Length: 771 July 17, 1992 10:37 Type: P Check: 3386 ..

1 MLPGLALLLL AAWTARALEV PTDGNAGLLA EPQIAMFCGR LNMHMNVQNG
 51 KWSDPSGTK TCIDTKEGIL QYCQEVYPEL QITNVVEANQ PVTIQNWCKR
 101 GRKQCKTHPH FVIPYRCLVG EFVSDALLVP DKCKFLHQR MDVCETHLHW
 151 HTVAKETCSE KSTNLHDYGM LLPCGIDKFR GVEFVCCPLA EESDNVDSAD
 201 AEEDSDVWW GGADTDYADG SEDKVVEVAE EEEVAEVEEE EADDEDDED
 251 GDEVEEEAEE PYEEATERTT SIATTTTTTT ESVEEVVREV CSEQAETGPC
 301 RAMISRWYFD VTEGKCAPFF YGGCGGNRNN FDTEEYCMAY CGSAMSQSLL
 351 KTTQEPLARD PVKLPTTAAS TPAVDKYLE TPGDENEHAH FQKAKERLEA
 401 KHRERMSQVM REWEEAERQA KNLPKADKKA VIQHFQEKVE SLEQEAANER
 451 QQLVETHMAR VEAMLNDRRR LALENYITAL QAVPPRPRHV FNMLKKYVRA
 501 EQKDRQHTLK HFEHVRMVDK KAAQIRSQV MTHLRVIYER MNQSLSLLYN
 551 VPAVAEEIQD EVDELLQKEQ NYSDDVLANM ISEPRISYGN DALMPSLTET
 601 KTTVELLPVN GEFSLDDLQP WHSFGADSVP ANTENEVEPV DARPAADRGL
 651 TTRPGSGLTN IKTEEISEV ^{Sueigh KM(670,117)112} ~~MDAEFRHDSG~~ YEVHHQKLVF FAEDVGSNKG
 701 AIIGLMVGGV ^{London U(713)I} ~~VIATVITL~~ ^{676 681 684} VMLKKKQYTS IHGVEVDA AVTPEERHLS
 751 KMQQNGYENP ^{T(714)STOP} TYKFFEQMQN *

⁷⁰
¹⁰² ~~MDAEFRHDSG~~ β -Amyloid domain

* human vs mouse amino acid differences

Ⓟ FAD mutations

Figure 16

Figure 17

18/47

8308
8386
8464
8542
8620
8698
8776
8854
8932
9010
9088
9166
9244
9322
9400
9478
9556
9634
9712
9790
9868
9946
10024
10102
10180
10258
10336
10414
XbaI (10493)
10492
10570
10648
10726
10804
10882
10960
11038
11116
11194
11272
11350
11428
EcoRV (11129)
11129
11207
11285
11363
11441
11519
11597
11675
11753
11831
11909
11987
12065
12143
12221
12299
12377
12455
12533
12611
12689
12767
12845
12923
13001
13079
13157
13235
13313
13391
13469
13547
13625
13703
13781
13859
13937
14015
14093
14171
14249
14327
14405
14483
14561
14639
14717
14795
14873
14951
15029
15107
15185
15263
15341
15419
15497
15575
15653
15731
15809
15887
15965
16043
16121
16199
16277
16355
16433
16511
16589
16667
16745
16823
16901
16979
17057
17135
17213
17291
17369
17447
17525
17603
17681
17759
17837
17915
17993
18071
18149
18227
18305
18383
18461
18539
18617
18695
18773
18851
18929
19007
19085
19163
19241
19319
19397
19475
19553
19631
19709
19787
19865
19943
20021
20099
20177
20255
20333
20411
20489
20567
20645
20723
20801
20879
20957
21035
21113
21191
21269
21347
21425
21503
21581
21659
21737
21815
21893
21971
22049
22127
22205
22283
22361
22439
22517
22595
22673
22751
22829
22907
22985
23063
23141
23219
23297
23375
23453
23531
23609
23687
23765
23843
23921
24000
24078
24156
24234
24312
24390
24468
24546
24624
24702
24780
24858
24936
25014
25092
25170
25248
25326
25404
25482
25560
25638
25716
25794
25872
25950
26028
26106
26184
26262
26340
26418
26496
26574
26652
26730
26808
26886
26964
27042
27120
27198
27276
27354
27432
27510
27588
27666
27744
27822
27900
27978
28056
28134
28212
28290
28368
28446
28524
28602
28680
28758
28836
28914
29000
29078
29156
29234
29312
29390
29468
29546
29624
29702
29780
29858
29936
30014
30092
30170
30248
30326
30404
30482
30560
30638
30716
30794
30872
30950
31028
31106
31184
31262
31340
31418
31496
31574
31652
31730
31808
31886
31964
32042
32120
32198
32276
32354
32432
32510
32588
32666
32744
32822
32900
32978
33056
33134
33212
33290
33368
33446
33524
33602
33680
33758
33836
33914
34000
34078
34156
34234
34312
34390
34468
34546
34624
34702
34780
34858
34936
35014
35092
35170
35248
35326
35404
35482
35560
35638
35716
35794
35872
35950
36028
36106
36184
36262
36340
36418
36496
36574
36652
36730
36808
36886
36964
37042
37120
37198
37276
37354
37432
37510
37588
37666
37744
37822
37900
37978
38056
38134
38212
38290
38368
38446
38524
38602
38680
38758
38836
38914
39000
39078
39156
39234
39312
39390
39468
39546
39624
39702
39780
39858
39936
40014
40092
40170
40248
40326
40404
40482
40560
40638
40716
40794
40872
40950
41028
41106
41184
41262
41340
41418
41496
41574
41652
41730
41808
41886
41964
42042
42120
42198
42276
42354
42432
42510
42588
42666
42744
42822
42900
42978
43056
43134
43212
43290
43368
43446
43524
43602
43680
43758
43836
43914
44000
44078
44156
44234
44312
44390
44468
44546
44624
44702
44780
44858
44936
45014
45092
45170
45248
45326
45404
45482
45560
45638
45716
45794
45872
45950
46028
46106
46184
46262
46340
46418
46496
46574
46652
46730
46808
46886
46964
47042
47120
47198
47276
47354
47432
47510
47588
47666
47744
47822
47900
47978
48056
48134
48212
48290
48368
48446
48524
48602
48680
48758
48836
48914
49000
49078
49156
49234
49312
49390
49468
49546
49624
49702
49780
49858
49936
50014
50092
50170
50248
50326
50404
50482
50560
50638
50716
50794
50872
50950
51028
51106
51184
51262
51340
51418
51496
51574
51652
51730
51808
51886
51964
52042
52120
52198
52276
52354
52432
52510
52588
52666
52744
52822
52900
52978
53056
53134
53212
53290
53368
53446
5352

Figure 18

2702 CTGCTTCCTTGCCCTAAGTATTCCTTTCTGATCACTATGCATTTTAAAGTAAACATTTTAAAGTATTCAGATGCTT
 2781 TAGAGAGATTTTTTTTCCATGACTGCATTTTACTGTACAGATTGCTGCTTCTGCTATATTTGTGATATAGGAATTAAGA

AIIIII (2866) HindIII (2921)

2860 GGATACACACGTTTGTTTCTCGTGCTGTTTTATGTGCACACATTAGGCATTGAGACTTCAAGCTTTTCTTTTTTTGT
 2939 CCACGTATCTTTGGGTCTTTGATAAAGAAAAGAATCCCTGTTTCATTGTAAGCACTTTTACGGGGCGGGTGGGGAGGGGT

EcoRI (3042)

3018 GCTCTGCTGGTCTTCAATTACCAAGAATTCTCCAAAACAATTTCTGCAGGATGATTGTACAGAATCATTGCTTATGAC
 3097 ATGATCGCTTTCTACACTGTATTACATAAAATAAATAAAATAAACCCTGGGCAAGACTTTTCTTTGAAGGATGACT
 3176 ACAGACATTAAATAATCGAAGTAATTTTGGGTGGGAGAAGAGGCAGATTCAATTTTCTTTAACCAGTCTGAAGTTTCA
 3255 TTTATGATACAAAAGAAGATGAAAATGGAAGTGCCAATATAAGGGGATGAGGAAGGCATGCC TGGACAAACC

3327 CTCTTTTAA GATGTGTCTT CAATTTGTAT AAAATGGTGT TTTCATGTAA ATAAATACAT TCTTGGAGGA
 3397 GCCACATGTG GCTGGTGTGA ATGATTCCAT AGTAACAATC TTGACCATTT ACTGACGTAC AGACCAGTGA
 3467 GAAGTCTTCG CATGTTGGGT ACCCACACCT GTGTGTCTT AATTGCAAGT CTGAGTAGGA AGTGGGGCC

AIIIII (3538)

3537 AACATGTGTC TCCAGTGCT GGGAAAATAT TTCATAGACC TAATTTACAG TCTTTACTTG ATCTAAACA
 3607 TTTTGCTGCC ATATTTTGGC CCTCAAGTTT GTCCCAAATG AGAGACAAAG GGAAAAGTTC CAGGGAAATA
 3677 AAAATTAAGA CAGCTGATTA TCTGTAAAGC ATGGTTTCTC ATCCTGAACG CTAATAACAT TTTGCAGGGA
 3747 ATAATTCCTT GTTGAAGGGA GTTGTCCTGA CCAGTGTAGG ATATTTATTT ATTTTATTTA TGTTTTTTGA
 3817 GACGGAGTCT CGCTCTGTCA CCCAGGCTGG AGTGCACTGG CACAATCTCG GCTCACTGCA AGCTCCGCCT
 3887 CCCGGGTTCG CGCCATCTC CTGCCTCAGC CTCCTGAATA GCTGGGACTC TAGGTGCCCC CCACCACGCC
 3957 CGGCTAATTT TTTGTATTTT TAGTAGAGAC GGGGTTTCAC CGTGTTAGCC AGGACAGTCT TGGTCTCCTG
 4027 ACCTCGTGAT CTGCCTGCCT CGGCCTCCCA AAGTGCTGAG ATTACAGGCG TGCAAGCCGC GCCCAGCCAG
 4097 TGCTCTCCTT TTAAAAGTAG CCCATTGGCT GGGCGCAGTG GCTCAGCCT GTAATCCCAG CACTTTGGGA
 4167 GGCTGAGGCG GGTGGATCAC GAGGTCAGGA GATCAAGAAT ATCCTGGCCA ATATGGTGAA ACCCATCTC
 4237 TACTAAAAAT ACAAAAAAAA AAAAAAAGGCG GGCATGGTGG CGGGCGCTTG TAGTCCCAGC
 4307 TACTCAGGAG GCTGAGGAG GAGAATGGTG TGCACCTGGG AGGCGGAGGT TGCAGTGAGC TGAGATCGCG
 4377 CCACTGCACT CCAGCCTGGG AGACAGAGCG AGACTCCGTC TCAATAAATA AATAAATAA TAAATAAAG
 4447 GAGGGCCTGG CACGAATGAC ATGCAGGGAA GGCAGTGAGC AGGTGGAGGT CCCTGTACTC GTTGTGGTGC

XbaI (41)

4517 CTTATCTACC AGGCGGTGA GTTGACGTCT TTGTGGACAG AATTCGAGCT CGGTACCCGG GGATCCTCTA

EcoRI (4556) BamHI (4577)

EcoRV (4601)

AIII (4595) AIII (4607)

HindIII (4628)

4587 GAGTCGAC CTTAAGGATATCCTTAAG GTCGACGGTA TCGATA AGCTTGGGCTTGAACATCGAGCGCCAGGGCTCC
 4662 GTAAAGCTACTAGAGCACAGGCGGTGCCCCAACGTCTGGGGCCTCTCCACTAATAACGGCTACTTCCAATTGATTGGA
 4741 CGCGCCATCTTGCTGCTTATGCATATTCAGCGGTGAAGTGAATATTCATGAACGAGGCCCGTCCCGTCCCTCCCTCC
 4820 TTCCCCCACCACCGGAACCCGCTCCGGAGGACCCGAAGGGCCCCGCCTTCATTACCGATGCGTAGGACAAACCATTTT
 4899 CCCGATGTGTGTGGGGGATACTAATGAGAGACTTTAGCTGAAAAATGAGCCTGAAGCTCCGAAGCTGAGTAAAAATGGC
 4978 CTAACCTTTATCTCCGTTCTGTAAGTCTCGGTTTGAGTGACGGGAAACCCGAAAGGAGGACGACAGGACCAGGACAT
 5057 TCTCTCTCTCTGTGCGTCAGAAAGAACACCCAACAGGGAGCCGGAGCCCTAGCGTCAACAACCTCCGCCCGCGCGCGC
 5136 TCCGTGTAGGCCGGTGCGGGCGGCCCGTAGCGCAAGGGAGGGCGGAAAGGAAGGGGCGGGACACAAGGGCGAATCTA
 5215 TAAAGGGCGTCACTCAGCCAGTTCTCTCTCAGAAGCGCCGAGAGCGGACCGGGACGGTTGGAGAAGAAGGTGGCTCC

EcoRI (5343)

5294 CGGAAGGGGGAGAGACAACTGCCGTAACCTCTGCCGTTTCAGGATC ATCGAATTC CTGCAGCCAAT ATG GGATCG
 1 Met GlySer
 5369 GCC ATT GAA CAA GAT GGA TTG CAC GCA GGT TCT CCG GCC GCT TGG GTG GAG AGG CTA TTC
 4 Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe
 5429 GGC TAT GAC TGG GCA CAA CAG ACA ATC GGC TGC TCT GAT GCC GCC GTG TTC CGG CTG TCA
 24 Gly Tyr Asp Trp Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser

NarI (5496)

5489 GCG CAG GGG CGC CCG GTT CTT TTT GTC AAG ACC GAC CTG TCC GGT GCC CTG AAT GAA CTG
 44 Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu
 5549 CAG GAC GAG GCA GCG CGG CTA TCG TGG CTG GCC ACG ACG GGC GTT CCT TGC GCA GCT GTG
 64 Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys Ala Ala Val
 5609 CTC GAC GTT GTC ACT GAA GCG GGA AGG GAC TGG CTG CTA TTG GGC GAA GTG CCG GGG CAG
 84 Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln
 5669 GAT CTC CTG TCA TCT CAC CTT GCT CCT GCC GAG AAA GTA TCC ATC ATG GCT GAT GCA ATG
 104 Asp Leu Leu Ser Ser His Leu Ala Pro Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met
 5729 CGG CGG CTG CAT ACG CTT GAT CCG GCT ACC TGC CCA TTC GAC CAC CAA GCG AAA CAT CGC
 124 Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg
 5789 ATC GAG CGA GCA CGT ACT CGG ATG GAA GCC GGT CIT GTC GAT CAG GAT GAT CTG GAC GAA
 144 Ile Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp Leu Asp Glu

[illegible]

12741 TTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAAACCCTCACTAAAGGGAACAAAAGCTG

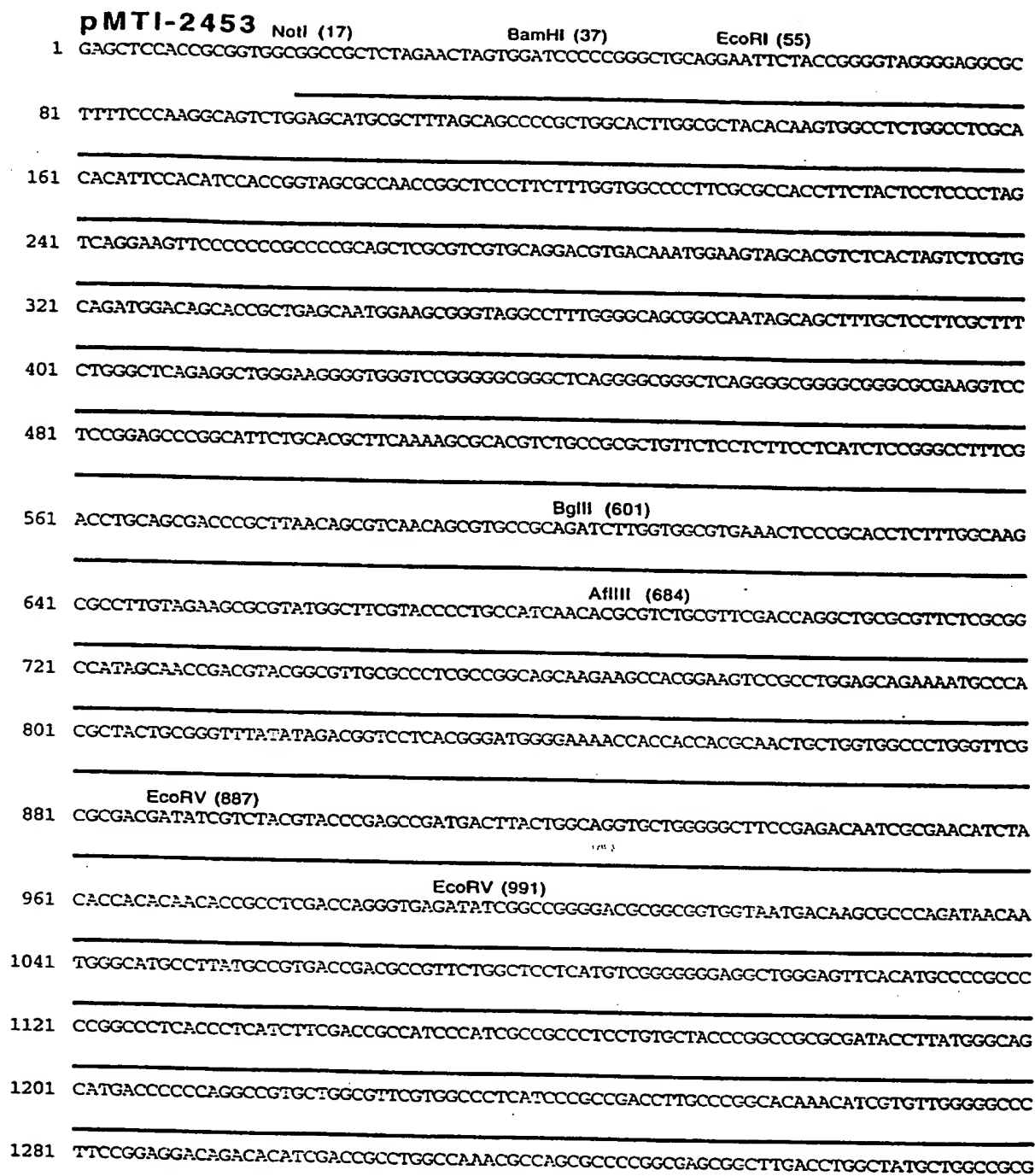


Figure 19

1361 ATTCGCL _GTTTACGGGCTGCTTGCCAATACGGTGGGTATCTGCAGGGC_ _GGGTGCTGGTGGGAGGATTGGGGACA

1441 GCTTTCGGGGACGGCCGTGCCGCCCCAGGGTGCCGAGCCCCAGAGCAACGCGGGCCACGACCCCATATCGGGGACACGT

1521 TATTTACCCGTGTTTCGGGCCCCGAGTTGCTGGCCCCAACGGCGACCTGTATAACGTGTTTGCCCTGGGCCTTGGACGTC

1601 TTGGCCAAACGCTCCGTCCCATGCACGTCTTTATCCTGGATTACGACCAATCGCCCGCCGGCTGCCGGGACGCCCTGCT

1681 GCAACTTACCTCCGGGATGGTCCAGACCCACGTCAACACCCAGGCTCCATACCGACGATCTGCGACCTGGCGCGCACGT

1761 TTGCCCCGGGAGATGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAAA

1841 AAGACAGAATAAAACGCACGGGTGTTGGGTGCTTTGTTTCATAAACGCGGGTTGGTCCCAGGGCTGGCACTCTGTGAT

1921 ACCCCACCGAGACCCCATTTGGGGCCAATACGCCCGGTTCTTCTTTTCCCCACCCCAACCCCCAAGTTGGGTGAAGG

2001 CCCAGGGCTCGCAGCCAACGTGCGGGCGGCAAGCCCGCCATAGCCACGGGCCCCGTGGGTTAGGGACGGGGTCCCCCATG

2081 GGGAAATGGTTTATGGTTCGTGGGGGTATTCTTTTGGGCGTTGCGTGGGGTCAGGTCCACGACTGGACTGAGCAGACAGA

2161 CCCATGGTTTTTGGATGGCTGGGCATGGACCGCATGTACTGGCGGACACGAACACCGGGCGTCTGTGGCTGCCAAACA

2241 CCCCCGACCCCCAAAAACCGCGCGGATTTCTGGCGCCCGGACGAATAAACCCTGACTACGGCATCTCTGCCCTT

2321 CTTCGCTGGTACGAGGAGCGCTTTTGTGTTTGTATTGGTCACCACGGCCGAGTTTCCGCGGACCCCGGCCAGGACCTGCA

2401 GAAATGATGATCTATTAAACAATAAAGATGTCCACTAAATGCAAGTTTTTCTCTGCATACCTTTGTTAAGAAGGGTGA

2481 GAACAGAGTACCTACATTTGAATGGAAGGATTGGAGCTACGGGGTGGCGGTGGGTGGGATTAGATAAATGCCTGCTC

2561 TTTACTGAAGGCTCTTTACTATTGCTTTATGATAATGTTTCATAGTTGGATATCATAATTTAAACAAGCAAAACCAATT

2641 AAGGGCCAGCTCATTCCTCCACTCATGATCTATAGATCTATAGATCTCTCGTGGGATCATGTTTCTCTTGTATGCCA

RNSDOCID: <WQ 9909150A1 1 >

NdeI (5525)
5521 TTTACATATGCTTTTAAAGATCGATGGGGATGCTTCATGTGAACGTGGGAGTTTCAGCTGCTTCTCTTGCCTAAGTATTCC
5601 TTTCCTGATCACTATGCATTTTAAAGTTAAACATTTTAAAGTATTTTCAGATGCTTTAGAGAGATTTTTTTTCCATGACTG

AflIII (5741)
5681 CATTTTACTGTACAGATTGCTGCTTCTGCTATATTTGTGATATAGGAATTAAGAGGATACACACGTTTGTTCCTCTCGTGC

HindIII (5796)
5761 CTGTTTTATGTGCACACATTAGGCATTGAGACTTCAAGCTTTTCTTTTTTGTCCACGTATCTTTGGGTCTTTGATAAAG

EcoRI
5841 AAAAGAATCCCTGTTTCATTGTAAGCACTTTTACGGGGCGGGTGGGAGGGGTGCTCTGCTGGTCTTCAATTACCAAGAAT
5921 TCTCCAAAACAATTTTCTGCAGGATGATTGTACAGAACTATTGCTTATGACATGATCGCTTTCTACACTGTATTACATAA
6001 ATAAATTAATAAAATAACCCGGCAAGACTTTTCTTTGAAGGATGACTACAGACATTAAATAATCGAAGTAATTTTGG
6081 GTGGGGAGAAGAGGCAGATTCAATTTTCTTTAACCAGTCTGAAGTTTCATTTATGATACAAAAGAAGTGAAGTGAAG
6161 TGGCAATATAAGGGGATGAGGAAGGCATGCTGACAAACCTTCTTTTAAGATGTGTCTTCAATTTGTATAAAATGGTG
6241 TTTTCATGTAAATAAATACATTCTTGGAGGAGCCACATTGTGCTGGTGTGAATGATTCCATAGTAACAATCTTGACCATT
6321 TACTGACGTACAGACCAGTGAGAAGTCTTCGCATGTTGGGTACCCACACCTGTGTGTCTTAATTGCAAGTCTGAGTAGG

AflIII (6413)
6401 AAGTTGGGGCCAACATGTGTCTCCAGTGCTGGGAAAATATTTTCATAGACCTAATTTACAGTCTTTACTTGATCTAAAAC
6481 ATTTTGTCTGCCATATTTTGGCCCTCAAGTTTGTCCCAAATGAGAGACAAAGGGAAAAGTTCCAGGGAAATAAAAAATTAAG
6561 ACAGCTGATTATCTGTAAAGCATGGTTTCTCATCTGAACGCTACTAACATTTTGCAGGGAATAATTCCTTGTTGAAGGG
6641 AGTTGTCTGACCAGTGTAGGATATTTATTTATTTATTTATTTATTTATTTTGTAGACGGAGTCTCGCTCTGTCAACCAGGCTG
6721 GAGTGCAGTGGCACAATCTCGGCTCACTGCAAGCTCCGCTCCCGGGTTACGCCATTCTCTCTGCTCAGCCTCCTGAAAT
6801 AGCTGGGACTCTAGGTGCCCCGCCACGCCCCGGCTAATTTTGTATTTTGTAGTAGAGACGGGGTTTACCGTGTTAGC
6881 CAGGACAGTCTTGGTCTCTCTGACCTCGTGATCTGCCTGCCTCGGCTCCCAAAGTGCTGAGATTACAGGGGTGCAAGCCG
6961 CGCCAGCCAGTGTCTCTCTTTTAAAAGTAGCCCATTTGGCTGGGCGCAGTGGCTCAGCCTGTAAATCCAGCACTTTGGG
7041 AGGCTGAGGCGGGTGGATCACGAGGTCAGGAGATCAAGAATATCCTGGCCAATATGGTGAAACCCCATCTCTACTAAAAA
7121 TACAAAAAAGAGGCGGGCATGGTGGCGGGCGCTTGTAGTCCCAGCTACTCAGGAGGCTGAGGCA
7201 GGAGAAATGGTGTGCACCTGGGAGGCGGAGGTTGCAGTGAGCTGAGATCGCGCCACTGCACTCCAGCCTGGGAGACAGAGC
7281 GAGACTCCGTCTCAATAAATAAATAAATAAATAAATAAAGAGGGCCTGGCACGAATGACATGCAGGGAAGGCAGTGAG

EcoRI (7431)
7361 CAGGTGGAGGTCCCTGTACTCGTTGTGGTGCCTTATCTACCAGGCGGTGAGTTGACGTCTTTGTGGACAGAATTCGAGC

BamHI (7452) HindIII (7491)
7441 TCGGTACCCGGGATCCTCTAGAGTCGACCTTAAGGTCGACGGTATCGATAAGCTTGGCCTTGAACATCGAGCGCCAGGG

7521 CTCCGTA...CTACTAGAGCACAGGCGGTGCCCCAACGTCTGGGGCCTCT...TAATAACGGCTACTTCCAATTGATT
7601 GGACGCGCCATCTTGCTGCCTTATGCATATTCAGCGGTGAACCTGAATATTCATGAACGAGGCCCCGTCCCGTCCCTCCCT
7681 CCTTCCCCCCCACCCCGGAACCCGCTCCGGAGGACCCGAAGGGCCCCGCCTTCATTACCGATGCGTAGGACAAACCATT
7761 TCCCGATGTGTGTGGGGGATACTAATGAGAGACTTTAGCTGAAAAATGAGCCTGAACCTCCGAAGCTGAGTAAAAATGGC
7841 CTAACCTTTATCTCCGTTCTGTAACTCTCGGTTTGAGTGACGGGAAACCCGAAAGGAGGACGACAGGACCAGGACATT
7921 CTCTCTCTCTGTGCGTCAAGAAACACCCAACAGGGAGCCGAGCCCTAGCGTCAACAACCTCCGCCGCGCGGCTC
8001 CGTGTAGGCCCGGTGCGGGCGCCCCGTAGCGCAAGGGAGGGCGGAAAGGAAGGGGCGGACACAAGGGCGAATCTATAA
8081 AGGGCGTCACTCAGCCAGTTCTCTCTCAGAAAGCGCGAGAGCGCGACCGGGACGGTTGGAGAAGAAGGTGGCTCCCGGA
EcoRI (8206)
8161 AGGGGGAGAGACAACTGCCGTAACCTCTGCGGTTTCAAGATCATCGAATTCCTGCAGCCAAATATGGGATCGGCCATTGAA
1 Met Gly Ser Ala Ile Glu
8241 CAAGATGGATTGCACGAGGTTCTCGGCCGCTTGGGTGGAGAGGCTATTTCGGCTATGACTGGGCACAACAGACAATCGG
7 Glu Asp Gly Leu His Ala Gly Ser Pro Ala Ala Thr Val Glu Arg Leu Phe Gly Tyr Asp Thr Ala Glu Glu Thr Ile Glu
NarI (8359)
8321 CTGCTCTGATGCCCGCGTGTTCGGGCTGTACGCGAGGGCGCCCGGTTCTTTTGTCAAGACCGACCTGTCCGGTGCCC
33 Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Glu Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala
8401 TGAATGAACTGCAGGACGAGGCGCGGCTATCGTGGCTGGCCACGACGGCGGTTCCTTGCGCAGCTGTGCTCGACGTT
60 eu Asn Glu Leu Glu Asp Glu Ala Ala Arg Leu Ser Thr Leu Ala Thr Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val
8481 GTCACCTGAAGCGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCTACCTTGCTCCTGC
87 Val Thr Glu Ala Gly Arg Asp Thr Leu Leu Leu Gly Glu Val Phe Gly Glu Asp Leu Leu Ser Ser His Leu Ala Pro Ala
8561 CGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAG
113 Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp His Glu N
8641 CGAAACATCGCATCGAGCGAGCACGTAATCGGATGGAAGCCCGTCTTGTGATCAGGATGATCTGGACGAAGAGCATCAG
140 Ile Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val Asp Glu Asp Asp Leu Asp Glu Glu His Glu N
NcoI (8792)
8721 GGGCTCGCGCCAGCCGAACCTGTTCCGCGAGGCTCAAGGCGCGCATGCCCCGACGGCGAGGATCTCGTGTGACCCATGGCGA
167 Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu Asp Leu Val Val Thr His Gly As
8801 TGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGCCCCGGCTGGGTGTGGCGGACC
193 Ala Cys Leu Pro Asn Ile Met Val Glu Asn Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp A
8881 GCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGCGCGGAATGGGCTGACCGCTTCTCTGCTGCTT
220 Arg Tyr Glu Asn Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Thr Ala Asp Arg Phe Leu Val Leu
8961 TACGGTATCGCGCTCCCGATTTCGAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGGGGATCAATTCTC
247 Tyr Gly Ile Ala Ala Pro Asp Ser Glu Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe • • •

12801 CTTATLCAATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAATTCCGCCCTA/AGTGAGTCGTATTACGCGCGCTCACTY3
12881 CCGTCGTTTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTTAATCGCCTTGACGACACATCCCCCTTTCCGCC
12961 AGCTGGCGTAATAGCGAAGAGGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGCGGAATGGGACGCGCC
13041 CTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCG
13121 CTCCTTTCCGCTTTCTTCCCTTCCCTTTCTCGCCACGTTCCGCCGGCTTTCCCGGTCAAGCTCTAAATCGGGGGCTCCCTTTA
13201 GGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAAATTGATTAGGGTGATGGTTACGTTAGTGGGACCTCTGTTCAAAACTGGAACAACAC
13281 CTGATAGACGGTTTTTCCGCCCTTTGACGTTGGAGTCCACGTTCTTAATAGTGGACTCTGTTCAAAACTGGAACAACAC
13361 TCAACCTATCTCGGCTCTATTCTTTGATTTTATAAGGGATTTTCCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATT
13441 TAACAAAAATTTAACGCGAATTTTAAACAAAATATTAAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCCGGGA
13521 ACCCTATTGTGTTTATTCTTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTTGATAAATGCTTCAATA
13601 ATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGCGCCCTTATTCCCTTTTTCGGGGCATTTTCGCTTCCTG
13681 TTTTGTCTACCCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCAGGAGTGGGTTACATCGAAGCTG
13761 GATCTCAACAGCGGTAAAGATCCTTGAGAGTTTTCGCCCCGAAGAAGCTTTTCCAATGATGAGCACTTTTAAAGTTCTGTCT
13841 ATGTGGCGGGTATTATCCCGTATTGACGCGGGGAAGAGCACTCGGTCGCCGCATACACTATTCTCAGAATGACTTTGG
13921 TTGAGTACTCAACAGTACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTTATGCAGTGCTGCCATAACCATG
14001 AGTGATAACACTGCGGCCAATCTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGG
14081 GGATCATGTAACCTCGCTTGTTCGTTGGGAACCGGAGCTGAATGAAGCCATACCAACGACGCGTGACACCAGCATGC
14161 CTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAAGTACTTACTCTAGCTTCCCGGCAACAATTAATAGAC
14241 TGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTCGGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGG
14321 AGCCGGTGAGCGTGGGTCTCGCGGTATCATTTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACA
14401 CGACGGGGAGTCAGGCAACTATGGATGAACGAATAGACAGATCGCTGAGATAGGTGCTCACTGATTAAAGCATTGGTAA
14481 CTGTCAAGCAAGTTTACTCATATATACTTTAGATTGATTAAAACTTCATTTTTTAATTTAAAGGATCTAGGTGAAGAT
14561 CCTTTTGTATAATCTCATGACCAAAATCCCTTAAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCA
14641 AAGGATCTTCTTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAAACAAAAACCACCGCTACCAGCGGTGGTT
14721 TGTTTGCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAAAATCTGTCTCT
14801 TCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCTGTGTAC
14881 CAGTGGCTGCTGCCAGTGGCGATAAGTGTGCTTTACCGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGG
14961 TCCGGCTGAACGGGGGTTCTGTGCACAGCCAGCTTGGAGCGAAGCAGCTACACCGAACTGAGATACCTACAGCGTGA
15041 GCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGC
15121 GCACGAGGGAGCTTCCAGGGGAAACGCTTGGTATCTTTATAGTCCCTGTCGGGTTTCCGCCACCTCTGACTTGAGCGTCA
15201 TTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGAAAAACGCCAGCAACGCGGCTTTTACGGTTCTTGCCCTTTTG
AIIII (15295)
15281 CTGGCCTTTTGTCTACATGTTCTTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTG
15361 ATACCGCTCGCCGCGAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAACCG
15441 CCTCTCCCGCGCGGTTGGCCGATTCTTAATGCAAGCTGGCAGCAGAGTTTCCCGACTGGAAGCGGGCAGTGAGCGCAA
15521 CGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAA
15601 TTGTGAGCGGATACAAATTTACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCTCACTAAAGG
15681 GAACAAAAGCTG

pMTI-2454 NotI (17) BamHI (37) EcoRI (55)

1 GAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCTACCGGGGTAGGGGAGGCGCT

82 TTTCCTCAAGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGGCACTTGGCGCTACACAAGTGGCCTCTGGCCTCGCACA

163 CATTCACATCCACCGGTAGCGCCAACCGGCTCCCTTCTTTGGTGGCCCCCTTCGCGCCACCTTCTACTCTCCCTAGTCA

244 GGAAGTTCCCCCGCGCCGCGAGCTCGCGTCTGTCAGGACGTGACAAATGGAAGTAGCAGTCTCACTAGTCTCGTGCAGA

325 TGGACAGCACCGCTGAGCAATGGAAGCGGGTAGGCCTTTGGGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTCTGGG

406 CTCAGAGGCTGGGAAGGGGTGGGTCCGGGGCGGGCTCAGGGGCGGGCTCAGGGGCGGGCGGGCGGAAGGTCTCCGGA

487 GCCCCGCATCTGCACGCTTCAAAGCGCAAGTCTGCGCGCTGTCTCTCTTCTCTCATCTCCGGGCCTTTCCGACCTGCA

BglII (601)

568 GCGACCCGCTTAACAGCGTCAACAGCGTCCCGCAGATCTTGGTGGCGTGAAACTCCCGCACCTCTTTGGCAAGCGCTTGT

AflIII (684)

649 AGAAGCGCGTATGGCTTCGTACCCCTGCCATCAACCGGCTCTGCGTTCCGACCAGGCTGCGCGTTCTCGCGGCCATAGCAA

730 CCGACGTACGGCGTTGCGCCCTCGCCGGCAGCAAGAAGCCACGGAAGTCCGCTGGAGCAGAAAATGCCACGCTACTGCG

EcoRV

811 GGTTTATATAGACGGTCTCTACGGCATGGGGAAAACCACCACCACGCAACTCTGGTGGCCCTGGGTTCGCGCGACGATAT

892 CGTCTACGTACCCGAGCCGATGACTTACTGGCAGGTGCTGGGGGCTTCCGAGACAATCGCGAACATCTACACCACACAACA

EcoRV (991)

973 CCGCCTCGACCAGGGTGAGATATCGGCCGGGACGCGCGGTGGTAATGACAAGCGCCCGAGATAACAATGGGCATGCCTTA

1054 TGCGGTGACCGACGCGGTTCTGGCTCCTCATGTGCGGGGGGAGGCTGGGAGTTCACATGCCCCGCCCCGGCCCTCACCTT

1135 CATCTTCGACCGCCATCCCATCGCGCCCTCCTGTGCTACCCGGCCGCGGATACCTTATGGGCAGCATGACCCCCAGGC

1216 CGTGTGGCGTTCTGTGGCCCTCATCCCGCGACCTTGCCCGGCACAAACATCGTGTGGGGGCCCTTCCGGAGGACAGACA

1297 CATCGACCGCCTGGCCAAACGCCAGCGCCCCGGCGAGCGGCTTGACCTGGCTATGCTGCGCGGATTGCGCGGTTTACGG

Figure 20

1378 GCTGCTAGCCAATACGGTGCGGTATCTGCAGGGCGGCGGGTCGTGGTGGGAGGATTGGGGACAGCTTTCGGGGACGGCCGT
AIII (1515)
1459 GCGCCCCAGGGTGCCGAGCCCCAGAGCAACGCGGGCCACGACCCCATATCGGGGACACGTTATTACCTGTTTCGGGC
AIII (1575)
1540 CCCCAGATTGCTGGCCCCAACGGCGACCTGTATAACGTGTTGCGCTGGGCCTTGACGTCTTGCCAAACGCCTCCGTCC
1621 CATGCACGTCTTTATCCTGGATTACGACCAATCGCCCGCCGGCTGCCGGGACGCCCTGCTGCAACTTACCTCCGGGATGGT
1702 CCAGACCCACGTACCCACCCAGGCTCCATACCGACGATCTGCGACCTGGCGCGCACGTTTGCCCGGGAGATGGGGGAGGC
1783 TAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGGT
1864 GTTGGGTCGTTTGTTCATAAACGCGGGGTTCCGTCCCAGGGCTGGCACTCTGTGATACCCACCGAGACCCCATTTGGGGC
1945 CAATACGCGCCGCTTTCTTCCTTTTCCCCACCCCAACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCAGCCAACGTGGG
NcoI (2076)
2026 GCGGCAAGCCCGCCATAGCCACGGGCCCCGTGGGTTAGGGACGGGTCCCCCATGGGAATGTTTATGGTTCGTGGGGT
NcoI (2162)
2107 TATTCTTTTGGGCGTTGCGTGGGGTCAGGTCCACGACTGGACTGAGCAGACAGACCAATGGTTTMTGGATGGCCTGGGCAT
2188 GGACCGCATGTACTGGCGGACACGAACACCGGGCGTCTGTGGTGCCAAACACCCCGACCCCAAAAACCACCGCGCG
NarI (2275)
2269 ATTTCTGGCGCCGCGGACGAACTAAACCTGACTACGGCATCTCTGCCCCCTTCTTCGCTGGTACGAGGAGCGCTTTTGT
2350 TGTATTGGTCACCAAGGCCGAGTTTCCGCGGGACCCCGCCAGGACCTGCAGAAATTGATGATCTATTAAACAATAAAGAT
2431 GTCCACTAAATGGAAGTTTTCCTGTCATACTTTGTTAAGAAGGGTGAGAACAGAGTACCTACATTTGAATGGAAGGA
2512 TTGGAGCTACGGGGGTGGGGGTGGGGTGGGATTAGATAAATGCCTGCTCTTTACTGAAGGCTCTTTACTATTGCTTTATGA
EcoRV (2609)
2593 TAATGTTTCATAGTTGGATATCATAATTAAACAAGCAAAACCAAATTAAGGGCCAGCTCATTCCTCCACTCATGATCTAT
BglII (2682)
BglII (2674)
2674 AGATCTATAGATCTCTCGTGGGATCATTTGTTTTCTCTTGATCCCACTTTGTGTTCTAAGTACTGTGGTTTCCAAATGTG

35/47

5467 CCCCTTAGCCAGTGTGATATTATTCTGTGGTTGTGACCCAATTAAGTCCTACTTTACATATGCTTTAAGAATCGATGGG NdeI (5525)
5548 GGATGCTTCATGTGAACGTGGGAGTTCAGCTGCTTCTCTGCTAAGTATTCCTTTCTGATCACTATGCATTTTAAAGTT
5629 AAACATTTTAAAGTATTTTCAGATGCTTTAGAGAGATTTTTCATGACTGCATTTTACTGTACAGATGCTGCTTCTGC

5710 TATATTTGTGATATAGGAATTAAGAGGATACACACGTTTGTTCCTGCTGCTGTTTATGTGCACACATTAGGCATTGAG AIII (5741)
HindIII (5796)
5791 ACTTCAAGCTTTTCTTTTGTGCCAGTATCTTTGGGTCTTTGATAAAGAAAAGAATCCCTGTTTCATTGTAAGCACTTTT

5872 ACGGGGCGGGTGGGAGGGGTGCTCTGCTGGTCTTCAATTACCAAGAATTCTCCAAAACAATTTTCTGCAGGATGATTGTA EcoRI (5917)
5953 CAGAATCATTTGCTTATGACATGATCGCTTTCTACACTGTATTACATAAAATAAAATAAAATAACCCCGGGCAAGACTT
6034 TTCTTTGAAGGATGACTACAGACATTAAATAATCGAAGTAATTTTGGGTGGGAGAAGAGGCAGATTCAATTTTCTTTAAAC
6115 CAGTCTGAAGTTTCATTTATGATACAAAAGAAGATGAAAATGGAAGTGGCAATATAAGGGGATGAGGAAGGCATGCCCTGGA
6196 CAAACCCTTCTTTTAAAGATGTGCTTCAATTTGTATAAAATGGTGTTCATGTAAATAAATACATTCTTGGAGGAGCCAC
6277 ATGTGCTGGTGTGAATGATTCCATAGTAACAATCTTGACCATTTACTGACGTACAGACCAGTGAGAAGTCTTCGCATGTT

6358 GGGTACCCACACCTGTTGTGTCTTAATTGCAAGTCTGAGTAGGAAGTTGGGGCCAACATGTGTCTCCAGTGCTGGGAAAA AIII (6413)
6439 TATTTATAGACCTAATTTACAGTCTTTACTTGATCTAAACATTTTGCTGCCATATTTTGGCCCTCAAGTTTGTCCCAA
6520 TGAGAGACAAAGGAAAAGTTCCAGGAAAATAAAATTAAGACAGCTGATTATCTGTAAAGCATGGTTTCTCATCCTGAAC
6601 GCTACTAACATTTTGCAGGGAATAATTCCTTGTTGAAGGGAGTTGCTCTGACCAGTGATAGGATATTTATTTATTTTATTA
6682 TGTTTTTTGAGACGGAGTCTCGCTCTGTACCCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAAGCTCCGCTC
6763 CCGGGTTCACGCCATCTCTCTGCTCAGCTCTCTGAATAGCTGGGACTCTAGGTGCCCCGCCACCAGCCCGGCTAATTTT
6844 TGTATTTTATAGTAGAGACGGGGTTTACCGTGTAGCCAGGACAGTCTTGGTCTCCTGACCTCGTGATCTGCCTGCCTCGG
6925 CCTCCCAAAGTGCTGAGATTACAGGCGTGAAGCCGCGCCAGCCAGTGTCTCTCTTTTAAAGTAGCCCATTTGGCTGGGC
7006 GCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGTGGATCACGAGGTCAGGAGATCAAGAATATCCT
7087 GGCCAATATGGTGAAACCCCATCTCTACTAAAAATACAAAAAAAAAAAAAAAAAAAAAGGCCGGCATGGTGGCGGGCG
7168 CTTGTAGTCCCAGCTACTCAGGAGGCTGAGGAGGAGAAATGGTGTGCACCTGGGAGGCGGAGGTTGCAGTGAGCTGAGATC
7249 GCGCCACTGCACTCCAGCCTGGGAGACAGAGCGAGACTCCGTCTCAATAAAATAAATAAATAAATAAAGGAGGGCCT
7330 GGCACGAATGACATGCAGGGAAGGCAGTGAGCAGGTGGAGGTCCCTGTACTCGTTGTGGTGCCTTATCTACCAGGCGGTTG

7411 AGTTGACGTCTTTGTGGACAGAATTCGAGCTCGGTACCCGGGATCCTCTAGAGTCGACCTTAAGGTGACGGTATCGATA EcoRI (7431) BamHI (7452) Hii
7492 AGCTTGGGCTTGAACATCGAGCGCCAGGGCTCCGTAAAGCTACTAGAGCACAGGCGGTGCCCAACGTCTGGGGCTCTC

7573 CACTAATAACGGCTACTTCCAATTGATTGGACGGCCATCTTGCTGCTTATGCATATTCAGCGGTGAACCTGAATATTCA
7654 TGAACGAGGCCCGTCCCGTCCCTCCCTCCCTTCCCCCACCCTCGGAACCCGCTCCGGAGGACCCGAAGGGCCCCGCTTCA
7735 TTACCGATGCGTAGGACAAACCATTTCCTCGATGTGTGTGGGGGATACTAATGAGAGACTTTAGCTGAAAAATGAGCCTG
7816 AACTCCGAAGCTGAGTAAAAATGGCTAACTTTATCTCCGTTCTGTAAGTCTCGGTTTGAGTGACGGGAAACCCGAAA
7897 GGAGGACGACAGGACCAGGACATTCTCTCTCTCTGTCTCGCTCAGAAAGAACACCCAACCCAGGGAGCCGGAGCCCTAGCGT
7978 CAACAACCTCCGCCCGCGCGCTCCCGTGTAGGCCGGTGCAGGGCGGCCCCGTAGCGCAAGGGAGGGCGGAAAGGAAGGGCGG
8059 GGACACAAGGGCGAATCTATAAAGGGCGTCACTCAGCCAGTTCTCTCTCAGAACGCGCCGAGAGCGGACCGGGACGGTTG
8140 GAGAAGAAGGTGGCTCCCGGAAGGGGAGAGACAACTGCCGTAACTCTGCCGTTTCAGGATCATCGAATCTCTGCAGCCA
EcoRI (8206)
8221 ATATGGGATCGGCCATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCCGCTATGACT
1▶ Met Gl ySer Al a l l e Gl u Gl n Asp Gl y Leu Hi s Al a Gl y Ser Pro Al a Al a T rp Val Gl u Arg Leu Phe Gl y Tyr Asp T
8302 GGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCGGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTGTCAAGA
27▶ r p Al a Gl n Gl n Thr l l e Gl y Cys Ser Asp Al a Al a Val Phe Arg Leu Ser Al a Gl n Gl y A rg P ro Val Leu Phe Val l Lys T
8383 CCGACCTGTCCCGTGCCCTGAATGAAGTGCAGGACGAGGCAGCGCGCTATCGTGGCTGGCCACGACGGGCGTTCTTGGC
54▶ hr Asp Leu Ser Gl y Al a Leu Asn Gl u Leu Gl n Asp Gl u Al a Al a Arg Leu Ser T rp Leu Al a Thr Thr Gl y Val Pro Cys A
8464 CAGCTGTGCTCGACGTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGCGAGGATCTCTGTGCAT
81▶ l a Al a Val Leu Asp Val Val l Thr Gl u Al a Gl y A rg Asp T rp Leu Leu Leu Gl y Gl u Val P ro Gl y Gl n Asp Leu Leu Ser S
8545 CTCACCTTGCTCTCTGCGGAGAAAGTATCCATCATGGCTGATGCAATGCCGCGGCTGCATACGCTTGATCCGGCTACCTGCC
108▶ er Hi s Leu Al a P ro Al a Gl u Lys Val Ser l l e Met Al a Asp Al a Met A rg Arg Leu Hi s Thr Leu Asp P ro Al a Thr Cys P
8626 CATTGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTA CTGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGG
135▶ r o Phe Asp Hi s Gl n Al a Lys Hi s Arg l l e Gl u Arg Al a Arg Thr A rg Met Gl u Al a Gl y Leu Val Asp Gl n Asp Asp Leu A
8707 ACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCC
162▶ sp Gl u Gl u Hi s Gl n Gl y Leu Al a P ro Al a Gl u Leu Phe Al a Arg Leu Lys Al a Arg Met P ro Asp Gl y Gl u Asp Leu Val V
NcoI (8792)
8788 TGACCCATGGCGATGCTTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGG
189▶ al Thr Hi s Gl y Asp Al a Cys Leu P ro Asn l l e Met Val Gl u Asn Gl y A rg Phe Ser Gl y Phe l l e Asp Cys Gl y A rg Leu G
8869 GTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGCGAATGGGCTGACCGCT
216▶ l y Val Al a Asp Arg Tyr Gl n Asp l l e Al a Leu Al a Thr A rg Asp l l e Al a Gl u Gl u Leu Gl y Gl y Gl u T rp Al a Asp Arg P
8950 TCCTCGTGCTTTACGGTATCGCCGCTCCCGATTGCGCAGCGCATCGCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGGGG
243▶ he Leu Val Leu Tyr Gl y l l e Al a Al a P ro Asp Ser Gl n Arg l l e Al a Phe Tyr Arg Leu Leu Asp Gl u Phe Phe . . .
9031 ATCAATTCTCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTGTGTTGCCCTCCCCCGTGC

38/47

13000 TCCCAA TTGCGCAGCCTGAAATGGGACGCGCCCTGTAGCGG...ATTAAACGCGCGGCGGTGTGGTGGTTACG
13081 CGCAGCGTGAACCGCTACACTTGCAGCGCCCTAGCGCCCGCTCCTTTCCGCTTCTTCCCTTCCCTTTCGCCACGTTCCGC
13162 GGCTTTCCCGTCAAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTAGTGCTTTACGGCACCTCGACCCCAAAAA
13243 CTTGATTAGGGTGATGGTTACAGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTT
13324 TTAAATAGTGGACTCTTGTTCCAACTGGAACAACACTCAACCCCTATCTCGGTCTATCTTTTGATTATAAGGGATTTTG
13405 CGGATTTCCGCCCTATTGGTTAAAAAATGAGCTGATTTAAACAAAAATTTAACCGGAATTTTAACAAAATATTACGCTTACA
13486 ATTTAGGTGGCACTTTTCCGGGAAATGTGCGCGGAACCCCTATTGTGTTTATTTTCTAAATACATTCAAATATGTATCCGC
13567 TCATGAGACAATAACCCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCC
13648 TTATTCCCTTTTTCGGGCATTTTGCCCTTCTGTTTGTCTCAGCCAGAAACGCTGGTGAAGTAAAGATGCTGAAGATC
13729 AGTTGGGTGCACGAGTGGGTTACATCGAAGCTGATCTCAACAGCGGTAAGATCCTTTGAGAGTTTTTCGCCCGAAGAAGCTT
13810 TTCCAAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCCGGCAAGAGCAACTCGGTC
13891 GCCGCAATCACTATCTCAGAAATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCATCTTAACGATGGCATGACAGTAA
13972 GAGAAATTATGACAGTGTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAG
14053 AGCTAACCGCTTTTTCACACAACATGGGGGATCATGTAACCTCGCTTGTATCGTTGGGAACCGGAGCTGAATGAAGCCATAC
14134 CAACCGACGAGCGTGACACCAAGATGCGCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTC
14215 TAGCTTCCCGGCACAAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCCTTCTGCGCTCGGCCCTTCCGGCTG
14296 GCTGGTTTATGCTGATAAATCTGAGCGCGGTGAGCGTGGGTCTCGCGGTATCATTTCCAGCACTGGGGCCAGATGGTAAGC
14377 CCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTG
14458 CCTCACTGATTAAAGCATTGGTAACCTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAACATTCATTTTAAAT
14539 TTAAGAGGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGT
14620 CAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTCGCGGTAAATCTGCTGCTTGCACAAACAAAAAAC
14701 CACCGCTACAGCGGTGGTTTGTTCGCGGATCAAGAGCTACCAACTCTTTTTCGGAAGGTAAGTGGCTTCAGCAGAGCGC
14782 AGATACCAAAATCTGTCTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGACCCGCTACATACCTCG
14863 CTCTGCTAATCCTGTTACAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTTGGACTCAAGACGATAGTTAC
14944 CGGATAAGGCGCAGCGGTGGGCTGAACGGGGGGTTTCGTGCACACAGCCAGCTTGGAGCGAAGACCTACACCGAACTGA
15025 GATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCGAGG
15106 TCGGACAGGAGAGCGCAAGAGGGAGCTTCCAGGGGGAAACGCTTGTATCTTTATAGTCTGTGCGGTTTTCGCCACCTCT
15187 GACTTGAGCGTCAATTTTGTGATGCTGTCAGGGGGGCGGAGCCTATGGAACCGCCAGCAACGCGCGCTTTTACGGT
AIIII (15295)
15268 TCCTGGCCTTTTGTCTGGCCTTTTGTCTCATATGTTCTTTCTCGGTTATCCCTGATTCTGTGGATAACCGTATTACCGCT
15349 TTGAGTGAGCTGATACCGCTCGCCGAGCCGACGACCGAGCGCAGCGAGTCACTGAGCGAGGAAGCGGAAGAGCGCCCAA
15430 TACGCAACCGCCTCTCCCGCGCGTTGGCCGATTCAATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAGCGGGCA
15511 GTGAGCGCAACGCAATTATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGT
15592 TGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAAACCTC
15673 ACTAAAGGGAACAAAAGCTG

pM1 2455 **Noll (17)** **PmeI (27)** **AscI (37)** **HindIII (62)**

1 GAGCTCCACCGCGGTGGCGGCCGCAAGTTTAAACATGGCGGCCCTTGTCGACGGTATCGATAAGCTTGATCTGATGGAAT

81 TAGAACTTGGCAAACAATACTGAGAATGAAGTGATGTGGAACAGAGGCTGCTGATCTCGTTCTTCAGGCTATGAAACT

161 GACACATTTGGAAACCACAGTACTTAGAACACAAAGTGGGAATCAAGAGAAAAACAATGATCCCACGAGAGATCTATAGA

241 TCTATAGATCATGAGTGGAGGAATGAGCTGGCCCTTAATTTGGTTTGTCTGTTTAAATTATGATATCCAACATGAAAC

321 ATTATCATAAAGCAATAGTAAAGAGCCTTCAGTAAAGAGCAGGCATTTATCTAATCCACCCCCACCCCCGTCAGC

401 TCCAATCCTTCCATTCAAAATGTAGGTACTCTGTTCTCACCCCTCTTAACAAAGTATGACAGGAAAAAACTTCCATTTTA

481 GTGGACATCTTTATTGTTTAAATAGATCATCAATTTCTGCAGGTCTGGCCGGGTCCCGCGGAAACTCCGCCGTGGTGAC

561 CAATACAAAACAAAAGCGCTCCTCGTACCAGCGAAGAAGGGGCAGAGATGCCGTAGTCAGGTTTAGTTCGTCCGGCGGCG

641 CCAGAAATCCGCGCGGTGGTTTGTGGGGTGGGGGTGTTTGGCAGCCACAGACGCCCGGTGTTGTTGTCGCGCCAGTAC

721 ATGCGGTCCATGCCAGGCCATCCAAAACCATGGGTCTGTCTGCTCAGTCCAGTCTGGAACCTGACCCACGCAACGCC

801 CAAAAGAATAACCCCCACGAACCATAAACCATTCCCCATGGGGGACCCCGTCCCTAACCCACGGGGCCCGTGGCTATGCC

881 GGGCTTGCCGCCCCGACGTGGCTGCGAGCCCTGGGCCTTACCCGAACCTGGGGGTGGGGTGGGGAAAAGGAAGAAAC

961 GCGGGCGTATTGGCCCCAATGGGGTCTCGGTGGGGTATCGACAGAGTGCCAGCCCTGGGACCGAACCCCGCGTTTATGAA

1041 CAAACGACCCAACACCCGTGCGTTTTATTCTGTCTTTTTATTGCCGTCATAGCGGGGTTCCTTCCGGTATTGTCTCCTT

1121 CCGTGTTCAGTTAGCCTCCCCATCTCCCGGCAACGTCGCGCCAGGTGCGAGATCGTCGGTATGGAGCCTGGGGTG

1201 GTGACGTGGGTCTGGACCATCCCGGAGGTAAGTTGCAGCAGGGCGTCCCGGACGCCGGGGCGATTGGTCGTAATCCAG

Figure 21

1281 GATAAAGACGTGCATGGGACGGAGGCGTTTGGCCAAGACGTCCAAGGCCAGGCAAACACGTTATACAGGTCGCCGTTGG
AIII (1337)

1361 GGGCCAGCAACTCGGGGGCCGAAACAGGGTAAATAACGTGTCCCGATATGGGGTCGTGGGCCCGCGTTGCTCTGGGGC
AIII (1397)

1441 TCGGCACCCTGGGGCGGCACGGCCGTCCCGAAAGCTGTCCCAATCCTCCACCAAGACCCGCCGCCCTGCAGATACCG

1521 CACCGTATTGGCAAGCAGCCCGTAAACGCGCGAATCGCGGCCAGCATAGCCAGGTCAAGCGCTCGCCGGGGGCGCTGGC

1601 GTTGGCCAGGCGGTTCGATGTGTCTGTCTCCGGAAGGGCCCCCAACACGATGTTTGTGCCGGCAAGGTCCGGCGGATG

1681 AGGGCCACGAACGCCAGCACGGCCCTGGGGGTTCATGCTGCCCATTAAGGTATCGCGCGGCCGGGTAGCACAGGAGGGCGGC

1761 GATGGGATGGCGGTTCGAAGATGAGGGTGAGGGCCGGGGCGGGGCATGTGAACTCCAGCCTCCCCCGACATGAGGAG

1841 CCAGAACGGCGTCGGTCACGGCATAAGGCATGCCATTGTTATCTGGGCGCTTGTTCATTACCACGGCCGCTCCCGGCC

EcoRV (1921)

1921 GATATCTCACCCCTGGTCGAGGCGGTGTTGTGTGGTGTAGATGTTCCGATTGTCTCGGAAGCCCCCAGCACCTGCCAGTA

EcoRV (2025)

2001 AGTCATCGGCTCGGGTACGTAGACGATATCGTCGCGGAACCCAGGGCCACCAGCAGTTGCGTGGTGGTGGTTTCCCCA

2081 TCCCGTGAGGACCGTCTATATAAACCCGAGTAGCGTGGGCATTTTCTGCTCCAGGCGGACTTCCGTGGCTTCTTGCTGC

AIII (2228)

2161 CGGCGAGGGCGCAACGCCGTACGTGGTTGCTATGGCCGCGAGAACGCGCAGCCTGGTTCGAACGCAGACGCGTGTGATG

BglII (2311)

2241 GCAGGGGTACGAAGCCATAACGCGCTTCTACAAGGCGCTTGCCAAAGAGGTGCGGAGTTTCAGCCACCAAGATCTCGG

2321 CACGCTGTTGACGCTGTTAAGCGGGTCGCTGCAGGTTCGAAAGGCCCGGAGATGAGGAAGAGGAGAACACGCGGCAGACG

2401 TGCGCTTTTGAAGCGTGCAGAATGCCGGGCTCCGGAGGACCTTCGCGCCCGCCCGCCCCCTGAGCCCGCCCCCTGAGCCCG

2481 CCCCCGACCCACCCCTTCCCAGCCTCTGAGCCCAGAAAGCGAAGGAGCAAAGCTGCTATTGGCCGCTGCCCCAAAGGCC

2561 TACCCGCTTCCATTGCTCAGCGGTGCTGTCCATCTGCACGAGACTAGTGAGACGTGCTACTTCCATTGTACGTCCTGC

2641 ACGACGCGAGCTGCGGGGCGGGGGGAACCTTCTGACTAGCGGAGGAGTAGAAGGTGGCGGAAGGGGCCACCAAAGAAG

42/47

5601 TAAGTATTCCTTTCTGATCACTATGCATTTTAAAGTTAAACATTTTAAAGTATTTAGATGCTTTAGAGAGATTTT

5681 TCCATGACTGCATTTTACTGTACAGATTGCTGCTTCTGCTATATTTGTGATATAGGAATTAAGAGGATACACACGTTTGT

AflIII (5751)

HindIII (5806)

5761 TTCTTCGTGCTGTTTATGTGCACACATTAGGCATTGAGACTTCAAGCTTTTCTTTTGTGCCAGTATCTTTGGGTC

5841 TTTGATAAAGAAAAGAATCCCTGTTTCATGTGAAGCACTTTACGGGGCGGGTGGGGAGGGGTGCTCTGCTGGTCTTCAAT

EcoRI (5927)

5921 TACCAAGAATTCTCCAAAACAATTTTCTGCAGGATGATTGTACAGAATCATTGCTTATGACATGATCGCTTTCTACACTG

6001 TATTACATAAATAAATAAATAAATAAATACCCCGGCAAGACTTTTCTTTGAAGGATGACTACAGACATTAAATAATCGAA

6081 GTAATTTTGGGTGGGAGAAGAGGCAGATTCAATTTTCTTTAACCAGTCTGAAGTTTCATTTATGATACAAAAGAAGATG

6161 AAAATGGAAGTGGCAATATAAGGGGATGAGGAAGGCATGCCTGGACAAACCCTTCTTTAAGATGTGTCTTCAATTTGTA

6241 TAAATGGTGTTTTCATGTAAATAAATACATTCTTGGAGGAGCCACATGTGTCTGGTGTGAATGATTCCATAGTAACAAT

6321 CTTGACCATTTACTGACGTACAGACCAGTGAGAAGTCTTCGCATGTTGGGTACCCACACCTGTTGTGTCTTAATTGCAAG

AflIII (6423)

6401 TCTGAGTAGGAAGTTGGGGCCAACATGTGTCTCCAGTGCTGGGAAAATATTTTCATAGACCTAATTTACAGTCTTTACTT

6481 GATCTAAAACATTTTGTGTCATATTTTGGCCCTCAAGTTTGTCCCAAATGAGAGACAAAGGGAAAAGTTCCAGGGAAAT

6561 AAAAATTAAGACAGCTGATTATCTGTAAAGCATGGTTTCTCATCTCTGAACGCTACTAACATTTTGCAGGGAATAATTCCT

6641 TGTGAAGGGAGTTGTCTGACCAGTGTAGGATATTTATTTATTTATTTATGTTTTTGTAGACGGAGTCTCGCTCTGTC

6721 ACCCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAAGCTCCGCCCTCCGGGTTCACGCCATCTCTCTGCTCAG

6801 CCTCTGAATAGCTGGGACTCTAGGTGCCGCCACCACGCCCGCTAATTTTGTATTTTAGTAGAGACGGGGTTTCA

6881 CCGTGTAGCCAGGACAGTCTTGGTCTCCTGACCTCGTGATCTGCCTGCCTCGGCCTCCCAAAGTGCTGAGATTACAGGC

6961 GTGCAAGCCGCGCCAGCCAGTGTCTCCTTTTAAAGTAGCCCATTTGGCTGGGCGCAGTGGCTCACGCCCTGTAATCCCA

7041 GCACCTTTGGGAGGCTGAGGCGGGTGGATCACGAGGTCAGGAGATCAAGAATATCCTGGCCAATATGGTGAAACCCCATCT

7121 CTACTAAAAATACAAAAAAGGCGCGGATGGTGGCGGGCGCTTGTAGTCCCAGCTACTCAGGA

7201 GGCTGAGGCAGGAGAATGGTGTGCACCTGGGAGGCGGAGGTTGCAGTGAGCTGAGATCGCGCCACTGCACTCCAGCCTGG

7281 GAGACAGAGCGAGACTCCGTCTCAATAAATAAATAAATAAATAAATAAAGGAGGGCCTGGCACGAATGACATGCAGGGA

7361 AGGCAGTGAGCAGGTGGAGGTCCCTGTACTCGTTGTGGTGCTTATCTACCAGGCGGTTGAGTTGACGCTTTGTGGACA

EcoRI (7441) BamHI (7462) HindIII (7501)

7441 GAATTCGAGCTCGGTACCGGGGATCCTCTAGAGTCGACCTTAAGGTCGACGGTATCGATAAGCTTGGGCTTGAACATCG

7521 AGCGCCAGGGCTCCGTAAAGCTACTAGAGCACAGGCGGTGCCCCAACGCTCTGGGGCTCTCCACTAATAACGGCTACTT

7601 CCAATTGATTGGACGCGCCATCTTGCTGCTTATGCATATTCAGCGGTGAAGTGAATATTCATGAACGAGGCGCGTCCC

7681 GTCCCTCTCTCTCTCCCCCACCCCGGAACCCGCTCCGGAGGACCCGAAGGCCCGCCTTTCATTACCGATGCGTAGGA
7761 CAAACCATTTCCTCCGATGTGTGTGGGGGATACTAATGAGAGACTTTAGCTGAAAAATGAGCCTGAACTCCGAAGCTGAG
7841 TAAAAATGGCCTAACTTTATCTCTCCGTCTCTGTAAGTCTCGGTTTGAGTGACCGGAAACCCGAAAGGAGGACGACAGGA
7921 CCAGGACATTCTCTCTCTCTCTGTCTCGCTCAGAAAGAACCCCAACCAGGGAGCCGGAGCCCTAGCGTCAACAACCTCCGCC
8001 GCGCGCGCTCCGTGTAGGCCGCTGCGGGCGGCCCGCTAGCGCAAGGGAGGGCGGGAAAGGAAGGGCGGGACACAAGGGC
8081 GAATCTATAAAGGGCGTCACTCAGCCAGTTCTCTCTCTCAGAAGCGCCGAGAGCGCGACCGGGACGGTTGGAGAAGAAGGT
8161 GGCTCCCGGAAGGGGAGAGACAACTGCGGTAACCTCTGCCGTTTCAGGATCATCGAATTCCTGCAGCCAATATGGGATC
EcoRI (8216)
1►Met Gl ySe
8241 GGCCATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCGCTTGGGTGGAGAGGCTATTCCGCTATGACTGGGCACAAC
3►rAlaIleGl uGlnAspGlyLeuHisAlaGlySerProAlaAlaTrpVal GluArgLeuPheGlyTyrAspTrpAlaGlnG
8321 AGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCTAGCGCAGGGGCGCCCGGTTCTTTTGTCAAGACCGACCTG
NarI (8369)
30►InThrIleGlyCysSerAspAlaAlaValPheArgLeuSerAlaGlnGlyArgProValLeuPheValLysThrAspLeu
8401 TCCGGTGCCCTGAATGAACTGCAGGACGAGGCGCGGCTATCGTGGCTGGCCACGACGGCGCTTCTTGCAGCTGT
57►SerGlyAlaLeuAsnGluLeuGlnAspGluAlaAlaArgLeuSerTrpLeuAlaThrThrGlyValProCysAlaAlaVa
8481 GCTCGACGTTGTCTACTGAAGCGGGAAGGGACTGGCTGCTATTGGCGAAGTGCCGGGCGAGGATCTCTGTCTATCTCACC
83►IleuAspValValThrGluAlaGlyArgAspTrpLeuLeuLeuGlyGluValProGlyGlnAspLeuLeuSerSerHisL
8561 TTGCTCTCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTC
110►euAlaProAlaGluLysValSerIleMetAlaAspAlaMetArgArgLeuHisThrLeuAspProAlaThrCysProPhe
8641 GACCACCAAGCGAAACATCGCATCGAGCGACGTACTCGGATGGAAGCCGGTCTTGTCTGATCAGGATGATCTGGACGA
137►AspHisGlnAlaLysHisArgIleGluArgAlaArgThrArgMetGluAlaGlyLeuValAspGlnAspAspLeuAspGly
8721 AGAGCATCAGGGGCTCGCGCCAGCCGAACGTGTTCCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCTCTGTGA
163►uGluHisGlnGlyLeuAlaProAlaGluLeuPheAlaArgLeuLysAlaArgMetProAspGlyGluAspLeuValValT
NcoI (8802)
8801 CCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGT
190►hrHisGlyAspAlaCysLeuProAsnIleMetValGluAsnGlyArgPheSerGlyPheIleAspCysGlyArgLeuGly
8881 GTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTT
217►ValAlaAspArgTyrGlnAspIleAlaLeuAlaThrArgAspIleAlaGluGluLeuGlyGlyuTrpAlaAspArgPh
8961 CCTCGTGCTTTACGGTATCGCCGCTCCCGATTCCGAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGGGG
243►eLeuValLeuTyrGlyIleAlaAlaProAspSerGlnArgIleAlaPheTyrArgLeuLeuAspGluPhePhe...
9041 ATCAATCTCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTGTGTTGCCCTCCCGGTG
9121 CCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCTCTTCTTAATAAAATGAGGAAATTCATCGCATGTCTGAGTAG

45/47

13201 GCTC TAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAA...CTTGATTAGGGTGATGGTTCACGTAGTG
13281 GGCCATGCCCCGTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGAGCTCTGTTCCTAACT
13361 GGAACAACACTCAACCCCTATCTCGGTCTATTCTTTTGATTATAAGGGATTTCGCCGATTTCGGCTTATGGTTAAAAAA
13441 TGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAAATATTAAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAA
13521 TGTGCGCGGAACCCCTATTGTGTTTATTTTCTAAATACATTCAAAATATGTATCCGCTCATGAGACAATAACCCCTGATAAA
13601 TGCCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCGGTGAAGATCAGTTGGGTGCACGAGTGGGTTA
13681 TGCCCTTCCTGTTTTTGTCTACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTA
13761 CATCGAACTGGATCTCAACAGCGGTAAAGATCCTTGAGAGTTTTTCGCCCGAAGAACGTTTCCAATGATGAGCACTTTTA
13841 AAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTGCGCCGATACACTATTCTCAG
13921 AATGACTTGGTTGAGTACTACCACTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAAATATGACAGTCTGC
14001 CATAACCATGAGTGATAACACTGCGGCCAACTTACTCTGACAAAGATCGGAGGACCGAAGGAGCTAACCGCTTTTTCG
14081 ACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGAGCTGAATGAAGCCATACCAAACGACGAGCGTGAC
14161 ACCACGATGCGCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACA
14241 ATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCGGCTGGCTGGTTTATTGCTG
14321 ATAAATCTGGAGCGGTGAGCGTGGGTCTCGCGGTATCATTGACAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTA
14401 GTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAATAGACAGATCGCTGAGATAGGTGCGCTCACTGATTAA
14481 GCATTGGTAACCTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTAAAACTTCATTTTAAATTTAAAGGATCT
14561 AGGTGAAGATCCTTTTGTATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCTAGAGCGTCAGACCCCGTA
14641 GAAAAGATCAAGGATCTTCTTGAGATCCTTTTCTGCGCGTAATCTGCTGCTTGCAAAACAAAAAACACCGCTACC
14721 AGCGGTGGTTTTGTGTCGGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAA
14801 ATACTGTCCTTCTAGTGTAGCGGTAGTTAGGCCACCACTTCAAGAATCTGTAGCACCGCCTACATACCTCGCTCTGCTA
14881 ATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
14961 GGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTCGAGCGAACGACCTACACCGAACTGAGATACC
15041 TACAGCGTGAGCTATGAGAAAGCGCCAGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCGAGGTCGGA
15121 ACAGGAGAGCGCACGAGGAGCTTCCAGGGGAAACGCTGCTATCTTTATAGTCTGTGCGGTTTCGCCACCTCTGACT
15201 TGAGCGTTCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGCCTTTTACGGTTCC
AIIIII (15305)
15281 TGGCCTTTTGTGCTGGCCTTTTGCTCACATGTTCTTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTT
15361 GAGTGAGCTGATAACGCTCGCCGCGAGCCGAACGACCGAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAAT
15441 ACGCAAACCGCCTCTCCCGCGCGTTGGCCGATTCAATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCA
15521 GTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTATTAGGCACCCAGGCTTTACACTTTATGCTTCCGGCTCGTATG
15601 TTGTGTGGAATTGTGAGCGGATAACAATTTCAACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCC
15681 TCACTAAAGGGAACAAAAGCTG

SphI (1292)
human APP 3' genomic/polyA

1	GCATGCCCTGG	ACAAACCCCTT	CTTTTAAGAT	GTGCTTCAA	TTTGTATAAA	ATGGTGTMTT	CATGTAAATA
71	AATACATTCT	TGGAGGAGCC	ACATTGTGCT	GGTGTGAATG	ATTCCATAGT	AACAATCTTG	ACCATTFACT
141	GACGTACAGA	CCAGTGAGAA	GTCTTCGCAT	GTGGGTACC	CACAOCTGTT	GTGTCTTAAT	TGCAAGTCTG
211	AGTAGGAAGT	TGGGGCCAAC	ATGTGTCTCC	CAGTGCTGGG	AAAAATATTC	ATAGAOCTAA	TTTACAGTCT
281	TTACTTGATC	TAAAACATTT	TGCTGCCATA	TTTGGCCCT	CAAGTTTGTC	CCAAATGAGA	GACAAAGGGA
351	AAAGTTCCAG	GGAAATAAAA	ATTAAGACAG	CTGATTATCT	GTAAAGCATG	GTTTCTCATC	CTGAACGCTA
421	CTAACATTTT	GCAGGGAATA	ATTCTGTGTT	GAAGGGAGTT	GTCCTGACCA	GTGTAGGATA	TTTATTATT
491	TTATTTATGT	TTTTTGAGAC	GGAGTCTCGC	TCTGTCACCC	AGGCTGGAGT	GCAGTGGCAC	AATCTGGCT
561	CACTGCAGC	TCCGCTCC	GGGTTCAGGC	CATTCTCTG	CCTCAGCTC	CTGAATAGCT	GGGACTCTAG
631	GTGCCCCCA	CCAGCCCCG	CTAATTTTTT	GTATTTTTAG	TAGAGACGGG	GTTTCACCGT	GTTAGCCAGG
701	ACAGTCTTGG	TCTCCTGACC	TCGTGATCTG	CCTGCCTCG	CCTCCCAAAG	TGCTGAGATT	ACAGGCGTGC
771	AAGCCGCGCC	CAGCCAGTGC	TCTCCTTTTA	AAAGTAGOCC	ATTGGCTGGG	CGCAGTGGCT	CAOGCTGTGA
841	ATCCAGCAC	TTTGGGAGGC	TGAGGOGGGT	GGATCAOGAG	GTCAGGAGAT	CAAGAATATC	CTGGCCAATA
911	TGGTGAAACC	CEATCTCTAC	TAAAAATACA	AAAAAAAAAA	AAAAAAAAAA	AAGGCCGGGC	ATGGTGGCGG
981	GCGCTTGTAG	TCCAGCTAC	TCAGGAGGCT	GAGGCAGGAG	AATGGTGTGC	AOCTGGGAGG	CGGAGGTTC
1051	AGTGAGCTGA	GATCGGCGCA	CTGCACTCCA	GCTGGGAGA	CAGAGOGAGA	CTCGTCTCA	ATAAATAAAT
1121	AAATAAATAA	ATAAAAGGAG	GGCTGGCAC	GAATGACATG	CAGGGAAGGC	AGTGAGCAGG	TGGAGGTCCC
1191	TGTACTCGTT	GTGTGCTT	ATCTACCAGG	CGGTTGAGTT	GACGTCTTTG	TGGACAGAAT	TCGAGCTGG

XbaI (1274)
 BamHI (1268) Sall (1280) SphI (1292)
 1261 TACCCGGGGA TCCTCTAGAG TCGACCTGCA GGCATGC

Figure 22

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/14507

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/00 C12N15/12 C12N15/10 C12N15/90 C12N5/10
C07K14/47 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>W0 95 14769 A (UNIV SOUTH FLORIDA) 1 June 1995</p> <p>see page 10, line 9 - page 13, line 14</p> <p style="text-align: center;">--- -/--</p>	<p>1,2,5,6, 8-11,16, 17, 19-24, 26,29, 31, 39-43, 45,52,54</p>



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

14 November 1997

Date of mailing of the international search report

08/12/1997

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

Hornig, H

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/14507

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 95 11968 A (ATHENA NEUROSCIENCES, INC.) 4 May 1995 see page 30, line 24 - page 39, line 22 ---	1,2,5,6, 8-11,16, 17, 19-24, 26,29, 31, 39-43, 45,52-54
X	HOWLAND D S ET AL: "MUTANT AND NATIVE HUMAN BETA-AMYLOID PRECURSOR PROTEINS IN TRANSGENIC MOUSE BRAIN" NEUROBIOLOGY OF AGING, vol. 16, no. 4, 1995, pages 685-699, XP000602061 see page 162, left-hand column, line 36 - line 41 ---	1,2,5,6, 8-11,16, 17,19, 20,24, 26,29, 31,39, 43,45, 52-54
A	GAMES D ET AL: "ALZHEIMER-TYPE NEUROPATHOLOGY IN TRANSGENIC MICE OVEREXPRESSING V717F BETA-AMYLOID PRECURSOR PROTEIN" NATURE, vol. 373, no. 6514, 9 February 1995, pages 523-527, XP000602050 cited in the application see the whole document ---	1-55
A	WO 94 23049 A (UNIV JOHNS HOPKINS) 13 October 1994 see the whole document ---	1-55
A	EP 0 653 154 A (HOECHST JAPAN) 17 May 1995 see the whole document ---	1-55
A	WO 95 20666 A (UNIV MINNESOTA ; HSIAO KAREN (US); BORCHELT DAVID R (US); SISODIA S) 3 August 1995 see the whole document ---	1-55
A	HASTY P ET AL: "INTRODUCTION OF A SUBTLE MUTATION INTO THE HOX-2.6 LOCUS IN EMBRYONIC STEM CELLS" NATURE, vol. 350, 21 March 1991, pages 243-246, XP002025690 cited in the application see the whole document ---	1-55

-/--

INTERNATIONAL SEARCH REPORT

In. ational Application No

PCT/US 97/14507

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>ASKEW G R ET AL: "SITE-DIRECTED POINT MUTATIONS IN EMBRYONIC STEM CELLS: A GENE-TARGETING TAG-AND-EXCHANGE STRATEGY". MOLECULAR AND CELLULAR BIOLOGY, vol. 13, no. 7, 1 July 1993, pages 4115-4124, XP000568152 cited in the application see the whole document</p>	1-55
A	<p>STACEY A ET AL: "USE OF DOUBLE-REPLACEMENT GENE TARGETING TO REPLACE THE MURINE X- LACTALBUMIN GENE WITH ITS HUMAN COUNTERPART IN EMBRYONIC STEM CELLS AND MICE" MOLECULAR AND CELLULAR BIOLOGY, vol. 14, no. 2, 1 February 1994, pages 1009-1016, XP000562809 cited in the application see the whole document</p>	1-55
A	<p>CAPECCHI M R: "ALTERING THE GENOME BY HOMOLOGOUS RECOMBINATION" SCIENCE, vol. 244, no. 4910, 16 June 1989, pages 1288-1292, XP000051938 see the whole document</p>	1-55
A	<p>K. DUFF ET AL.: "Insertion of a pathogenic mutation into a yeast artificial chromosome containing the human amyloid precursor protein gene" GENE THERAPY, vol. 1, no. 1, January 1994. MACMILLAN PRESS, UK, pages 70-75, XP002046823 see the whole document</p>	1-55
P,X	<p>WO 96 34097 A (CEPHALON INC ;SCOTT RICHARD W (US); REAUME ANDREW G (US); TRUSKO S) 31 October 1996 see page 21, line 12 - line 30 see page 22, line 1 - line 8</p>	1,2
P,X	<p>WO 96 40895 A (ATHENA NEUROSCIENCES INC) 19 December 1996 see page 29, line 27 - line 32</p>	1-13, 16-54

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 97/14507

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9514769 A	01-06-95	NONE	
WO 9511968 A	04-05-95	US 5604102 A US 5612486 A AU 8079894 A AU 8080994 A CA 2174429 A CA 2174632 A EP 0736106 A EP 0730643 A WO 9511994 A	18-02-97 18-03-97 22-05-95 22-05-95 04-05-95 04-05-95 09-10-96 11-09-96 04-05-95
WO 9423049 A	13-10-94	NONE	
EP 0653154 A	17-05-95	JP 7132033 A CA 2135595 A	23-05-95 13-05-95
WO 9520666 A	03-08-95	AU 1909695 A EP 0742831 A	15-08-95 20-11-96
WO 9634097 A	31-10-96	US 5586226 A	17-12-96
WO 9640895 A	19-12-96	AU 6264996 A	30-12-96

THIS PAGE BLANK (USPTO)